

密级: _____

浙江大学

硕 士 学 位 论 文



论文题目: 基于表面肌电信号的手势分类
——改进的深度森林方法

作者姓名: 刘方昊

指导教师: 张国川

学科 (专业): 计算机技术

所在学院: 计算机科学与技术学院

提交日期: 二〇二〇年十二月二十日

A Dissertation Submitted to
Zhejiang University for the Degree of
Master of Engineering



TITLE: Gesture classification using sEMG signals
——An improved deep forest algorithm

Author: Fanghao Liu

Supervisor: Guochuan Zhang

Subject: Computer Techonology

College: Computer Science and Techonology

Submitted Date: 2020-12-20

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的
研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发
表或撰写过的研究成果，也不包含为获得 浙江大学 或其他教育机构的学位或
证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文
中作了明确的说明并表示谢意。

学位论文作者签名: 签字日期: 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权浙江大学可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名： 导师签名：

签字日期： 年 月 日 签字日期： 年 月 日

摘要

随着智能化产品不断融入人类生活，手势识别在人机交互中的重要性日益凸显。表面肌电信号 (Surface Electromyography, sEMG) 是伴随肌肉活动产生的一种重要的生物电信号，可实时、准确地反映肌肉的活动和状态。基于表面肌电信号的手势识别在肌电假肢等应用场景都具有重要意义。

由于基于表面肌电信号的手势分类问题通常具有数据维数高、样本数量少、目标类别多的特点，因此直接使用传统的机器学习算法或深度学习算法都难以很好地解决问题。目前已有的研究多依赖专家知识或复杂的信号处理进行特征提取以提升分类精度。Zhou 和 Feng 提出的深度森林方法很适合解决小样本高维数据上的分类问题。在此基础上，本文提出了一种改进的深度森林算法来解决基于表面肌电信号的手势分类问题。核心思想包括三方面：(1) 在每层中对每个类别训练一个随机森林，每个随机森林重点学习一个类别与其它所有类别的差异，从而充分提取特征信息；(2) 借鉴提升学习的思路，在同一层中序列化地构造两个随机森林，训练第二个森林使用的特征由第一个森林的得到的特征重要性确定，从而使得提取的特征更加互补；(3) 在级联结构中，根据特征重要性对前一层的特征进行特征选择后传递到下一层，使得模型能够提取到更高阶的特征，并以更快的速度收敛。

实验结果表明，本文提出的模型在 UCI 手势识别数据集上具有很好的分类效果，取得了 99.38% 的平均分类精度。

关键词：表面肌电，手势识别，深度森林

Abstract

With the continuous integration of intelligent products into human life, gesture recognition plays an increasingly important role in human-computer interaction. Surface Electromyography (sEMG) is an important bioelectrical signal generated with muscle activity, which can reflect the activity and state of muscles in real time and accurately. Gesture recognition based on surface EMG signal is of great significance in electronic prosthesis and other application scenarios.

Since gesture classification based on surface EMG signal is usually characterized by high data dimension, small sample size and large categories, it is difficult to solve the problem by direct use of traditional machine learning algorithms or deep learning algorithms. At present, most of the existing researches rely on expert knowledge or complex signal processing for feature extraction to improve classification accuracy. The deep forest method proposed by Zhou and Feng is suitable for solving the classification problem of small sample data with high dimension. Based on this method, an improved deep forest algorithm is proposed to solve the gesture classification problem based on surface EMG signal. The core idea includes three aspects : (1) train a random forest for each category in each layer, and each random forest focuses on learning the difference between one category and all other categories, so as to fully extract feature information; (2) two random forests are constructed sequentially in the same layer by referring to the idea of Boosting, and the features used by the second forest are determined by the importance of the features obtained from the first forest, so as to make the extracted features more complementary; (3) in a cascading structure, the features of the previous layer are selected based on the feature importance and then transferred to the next layer, so that the model can extract features of higher order and converge at a faster speed.

The experimental results show that the improved deep forest algorithm has a good classification effect on the UCI gesture recognition data set and achieves an

average classification accuracy of 99.38%.

Keywords: Gesture Classification, Surface Electromyography, Deep Forest

目录

摘要	I
Abstract	II
第 1 章 绪论	1
1.1 研究背景	1
1.2 研究现状	2
1.3 本文主要工作	4
1.4 章节安排	5
第 2 章 相关算法介绍	6
2.1 随机森林	6
2.2 深度神经网络	8
2.3 深度森林	9
2.3.1 级联森林 (Cascade Forest)	9
2.3.2 多粒度扫描 (Multi-Grained Scanning)	12
2.3.3 深度森林 (gcForest)	14
2.4 局部二值模式	16
2.5 本章小结	18
第 3 章 改进的深度森林算法	19
3.1 带类别权重的随机森林	19
3.2 序列化训练的随机森林	22
3.3 层级间特征选择机制	25
3.4 改进方法的整体框架	27
3.5 本章小结	29
第 4 章 实验设计与结果分析	30
4.1 数据集简介	30
4.2 基于 sEMG 进行手势分类问题的实验设计	32
4.2.1 使用原始信号作为输入	32
4.2.2 使用提取的特征作为输入	35
4.3 基于其它数据集的实验设计	41

4.4 本章小结	45
第 5 章 总结与展望	47
参考文献	49
攻读硕士学位期间的主要研究成果	53
致谢	54

插图

2.1	随机森林示意图	7
2.2	深度神经网络示意图	8
2.3	级联森林结构示意图 ^[18]	10
2.4	类别分布向量计算过程示意图 ^[18]	11
2.5	多粒度扫描流程示意图 ^[18]	13
2.6	深度森林完整流程图 ^[18]	14
2.7	2DLBP 示意图.....	16
2.8	1DLBP 示意图.....	17
3.1	单层中带类别权重的森林示意图	20
3.2	单层中序列化训练的随机森林示意图	23
3.3	层级间特征选择机制示意图	26
3.4	改进方法的整体框架	27
4.1	六个手势示意图 ^[8]	30
4.2	六个手势的 2 通道表面肌电信号示意图.....	31
4.3	原始信号和对应的功率频谱图	36
4.4	两个不同的信号对应相同的 LBP 值	37
4.5	从 sEMG 原始信号中提取 LBP ^[7]	38

表格

2.1	gcForest 模型中的超参数.....	15
3.1	改进后模型中的超参数.....	28
4.1	各分类器的超参数设置.....	33
4.2	各分类器实验结果比较.....	34
4.3	表面肌电信号的时频域特征	35
4.4	各分类器实验结果比较——手势分类.....	39
4.5	与之前研究结果的比较——手势分类.....	40
4.6	各分类器实验结果比较——音乐分类.....	42
4.7	各分类器实验结果比较——情绪识别.....	43
4.8	各分类器实验结果比较——新闻分类.....	44
4.9	各分类器实验结果比较——DNA 分类.....	45

第 1 章 绪论

本章介绍了基于表面肌电信号的手势识别问题的研究背景、研究意义、现有方法存在的问题以及本文采纳的研究方法。随后介绍了研究现状，列举了近年来关于表面肌电信号的手势识别的主要研究方法及结果，并简要陈述了本文的主要工作。最后给出了本文的章节组织架构。

1.1 研究背景

表面肌电信号 (surface electromyography, sEMG) 是由电极周围肌纤维产生的运动单元电位在空间和时间上的累积。这些信号取决于运动相关的肌肉收缩过程中的解剖和生理特性，能够为人们提供与神经肌肉活动相关的重要信息^[1]。表面肌电信号作为一种生物信号，在人机交互装置的研制中发挥着重要的作用。近年来，肌电控制技术受到了生物医学研究者的广泛关注。表面肌电信号和运动之间的相关性使肌电信号成为控制假肢的最强大的信号源之一^[2]。

长期以来，疾病、自然灾害、交通事故以及各种意外伤害造成了成千上万人的肢体残疾，仅在我国就大约有 226 万的肢体截肢者^[3]。为截肢者安装假肢是恢复其肢体功能的主要手段。研究和开发先进的假肢以及假肢控制系统，改善和提高假肢的操控性能，为众多截肢者提供智能化、多功能的假肢，能够极大地提高他们的生活品质。传统的假肢存在功能单一、操控缓慢、动作笨拙和维护困难等问题。

基于表面肌电信号开发的肌电假肢可以通过对残留肢体肌肉表面采集的 sEMG 信号进行解码，从而得知截肢者想要做的肢体动作类型，实现高自由度的假肢控制。虽然脑电也可以用于识别截肢者的动作意图，但就舒适度而言，佩戴含有肌电电极的手部装置比佩戴含有脑电电极的头部装置要更加舒适，因此肌电假肢得到了研究者的广泛关注。

基于表面肌电信号的手势识别是利用肌电控制假肢需要解决的关键问题。建立一个能基于表面肌电信号将手势准确分类的模型对开发更加精准的肌电假肢

系统至关重要。一方面，我们希望可以提升主要手势的分类精度，另一方面，我们也希望可以对更多的手势进行分类，从而提升机电假肢的自由度。

由于表面肌电信号是一种生物信号，通常具有样本维度高（因为采样率通常较高）、样本量较小（因为提供样本的受试者有限）的特点，因此目前的经典机器学习算法与深度学习算法都无法较好的解决基于表面肌电信号的手势分类问题。支持向量机、随机森林、K 近邻等经典机器学习算法的模型复杂度有限，且从小样本高维特征中提取特征并进行特征转换的能力不够强，往往难以取得较好的表现。深度学习方法虽然具有较高的模型复杂度和较强的特征提取和转换能力，但深度学习往往需要较多的数据，在小样本的数据集上容易过拟合^[29]。

Zhou 和 Feng^[18] 提出了一种结合了随机森林和神经网络结构的算法——深度森林 (gcForest)，实验证明该模型在多场景的多个分类任务中具有和深度学习方法相当甚至更优的表现，且能较好地适用于样本较小的情况，因此比较适合解决基于表面肌电信号的手势识别的问题。本文基于深度森林的方法做了一些改进，并将改进后的方法用于基于表面肌电信号的手势识别问题，取得了较好的分类效果。

1.2 研究现状

基于表面肌电信号 (sEMG) 的手势分类和识别技术发展到目前为止已经有很多相关研究。

Sapsanis 等人^[4] 提出了一种基于表面肌电信号的手势识别方法。为了验证该方法的可行性，他们选取了 5 名年龄相仿的健康受试者，并收集了他们的 6 个基本手势的双通道肌电信号。首先利用经验模态分解 (EMD) 将表面肌电信号分解成本征模函数 (IMF)，并对各 IMF 和原始信号提取常用的八个时域和频域特征，包括 Willison 振幅、方差、过零率等，最后利用简单的线性分类器对表面肌电信号进行分类，取得了 89.2% 的平均分类精度。

Iqbal 等人^[5] 先提取短时重叠子帧，形成子帧矩阵，接着对该矩阵进行奇异值分解来提取奇异值和主成分，在提取奇异值的基础上，提出了前 5 个主成分的一些统计参数作为特征，最后采用 k-NN 进行手势识别。文中指出，所提出的方法能够以较高的准确率和较低的执行时间对 6 种基本手势进行分类。

Akben^[6] 采用滤波和直方图计算的方式提取特征, 并利用一致性相关法来计算直方图之间的相关关系。最后, 用一致性相关系数作为分类器的输入, 并采用级联结构化分类器对手势进行分类。平均精度为 94.72%。

Tabatabaei^[7] 将一维局部二值模式 (1DLBP) 方法应用于表面肌电信号手势分类, 并通过特征选择过滤掉不重要的模式和噪声, 随后使用支持向量机、logistic 回归、判别分析和 KNN 等多种经典的机器学习分类器来处理提取出的特征, 结果表明支持向量机的表现最佳, 平均精度为 92.9%。

Nishad 等人^[8] 引入了一种基于可调 q 小波变换 (TQWT) 的滤波器组 (FB), 首先利用该滤波器组 TQWT-FB 对表面肌电信号进行交叉协方差分解, 随后利用计算 Kraskov 熵的方式提取特征, 最后使用 KNN 分类器来进行手势分类, 平均精度为 98.55%。

Subasi^[9] 等人采用多尺度主成分分析 (MSPCA) 方法去噪, 采用小波包分解 (WPD) 方法进行特征提取, 并采用不同分类器评估效果, 结果表明旋转森林分类器效果最佳, 取得了 93.44% 的平均精度。

Sikder^[10] 等人采用自回归方法得到原始表面肌电信号的功率谱密度估计, 并把功率谱密度作为多通道卷积网络的输入进行手势分类, 取得了不错的结果, 平均分类精度为 96.3%。

Ayaz^[11] 等人采用短时傅里叶变换的方法得到原始信号的时频图像, 然后采用基于局部感受野的极限学习机 (ELM-LRF) 进行手势分类, 最终平均分类精度为 94.12%。

Song^[12] 等人使用卷积神经网络 (CNN) 来自动提取肌电信号特征, 然后使用 dueling deep Q-learning 来学习一种分类策略, 该策略能够选择最有用的特征子集, 并过滤不相关或冗余的特征。结果表明, 该方法优于多层感知器 (MLP), 可以使用较少的特征来达到较高的分类精度。

Rabin^[13] 等人采用短时傅里叶变换进行特征提取, 并采用 diffusion maps (DM) 进行降维。作者提出了一种仿射变换, 用于对两个被试者的特征空间对齐, 最后采用 k 近邻法 (K -nearest neighbour, KNN) 进行分类。研究结果表明, 在训练数据有限的情况下, DM 优于 PCA, 且提出的方法使得基于某一被试者的数据集训练的模型可以对新被试者的数据集进行鲁棒分类。

Kakoty^[31] 等人对表面肌电信号进行最大等长收缩标准化, 并采用基于离散小波变换系数的主成分分析, 提取出表面肌电信号的特征向量, 最后通过基于径向基核函数的支持向量机进行分类, 取得了 97.5% 的平均分类精度。

Yavuz^[32] 等人利用倒频谱分析技术, 通过计算倒频谱系数有效地提取了表面肌电信号的时域特征, 随后将提取的由 MFCC 组成的特征向量输入到广义回归神经网络 (GRNN) 中去, 取得了 99.34% 的平均分类精度。

上述研究大多数先基于某些方法提取特征, 如 EMD、SVD、DWT、WPD、1DLBP 和 TQWT, 然后采用常用分类方法如 k-NN、MLP、RF、LDA 和 SVM 得出最终结果。还有部分研究利用深度学习来对经过降维处理的数据进行学习得到分类结果。前者依赖较多的专家知识, 而后者则有耗时长、易过拟合的问题, 本文针对这些问题提出了一个新方法, 即利用改进的深度森林算法来解决手势分类问题。

1.3 本文主要工作

本文提出了一种改进的深度森林方法来解决基于表面肌电信号的手势分类问题。具体包括以下几方面的工作:

(1) 为了使得深度森林方法更适用于解决基于表面肌电信号的手势分类这样的小样本、高维数据上的多分类问题, 本文基于深度森林 (gcForest) 的不足做了三方面改进, 包括: a、每个类别训练一个随机森林; b、借鉴提升方法的思路在同一层中序列化训练随机森林; c、在层级之间传递特征的时候引入特征选择机制。

(2) 将上述改进方法应用在 UCI 机器学习库的表面肌电信号手势分类数据集上, 比较了本文的改进方法与经典的机器学习方法和深度学习方法的实验结果, 表明了基于表面肌电信号的手势分类问题中本文提出的方法相比传统机器学习方法和深度学习方法具有更好的表现。此外, 我们还通过实验比较了改进方法与原深度森林 gcForest 模型的表现, 表明了改进后的方法比原深度森林方法的表现有所提升。最后我们将本文提出的改进方法的实验结果与之前已有的研究结果进行了对比, 验证了该方法在基于表面肌电信号的手势分类问题中具有一定的优越性。

(3) 为说明改进后的方法不仅适用于基于表面肌电信号的手势分类问题，还具有一定普适性，我们将本文提出的方法应用于其它几个公开数据集，并与经典分类器结果进行比较，验证了模型的效果，表明了该方法在多场景的分类任务中均有较好表现，且尤其适用于高维小样本数据上的多分类问题。

1.4 章节安排

本文分为五章：

第一章介绍基于 sEMG 的手势识别问题的研究背景以及该问题的研究意义，综述现有的研究结果，陈述本文主要工作，并给出论文的章节安排。

第二章简述了本文提出的模型所涉及的基本算法，包括随机森林和深度神经网络，并基于此阐述了深度森林（Deep Forest）的基本思想以及算法流程。随后介绍了局部二值模式（Local Binary Pattern）方法，该方法在本文中用于 sEMG 的特征提取。

第三章给出了本文提出的改进算法。首先提出原深度森林方法的不足，并针对不足提出改进措施，接着建立了模型的整体框架，并阐述算法流程。

第四章为实验部分，包括基于表面肌电信号的手势分类实验和基于其它公开数据集的实验。在各实验中，我们给出实验数据集简介，特征提取方法，所用模型及参数设置，以及最终实验结果的对比分析。

第五章对本文研究工作进行了总结，针对不足提出了可能的改进方向，并对接下来可开展的工作进行了展望。

第2章 相关算法介绍

本章提及的算法包括两部分，第一部分是本文的改进方法所涉及的算法，第二部分是对表面肌电信号提取特征所涉及的算法。第一部分简述了随机森林、深度神经网络的原理和算法流程，基于此给出了深度森林的详细介绍；第二部分介绍了用于表面肌电信号时域特征提取的局部二值模式方法。

2.1 随机森林

集成学习 (Ensemble Learning) 通过构建并结合多个学习器来完成学习任务，有时也被称为多分类器系统。集成学习大致可以分为两类：Boosting 和 Bagging (bootstrap aggregating)。Boosting 通过在训练新学习器时更关注前一学习器错误分类的实例来序列化地训练多个弱学习器，增量地构建集成模型。Bagging 使用随机抽取的子训练集训练各个弱学习器，然后让各个学习器投票得出最终结果，各学习器投票时具有相同的权重。

随机森林 (Random Forest, 简称 RF) 是 Bagging 方法的一个拓展变体，由 Breiman^[14] 于 2001 年提出。随机森林在以 CART (classification and regression tree) 为基学习器构建 Bagging 模型的基础上，进一步在各决策树分类器的训练过程中引入随机特征选择。具体来说，传统决策树在选择划分属性时是在当前结点的特征集合中选择一个最优特征，而在随机森林中，基决策树分类器在进行训练时先从当前结点的特征集合中随机抽取一个特征子集，然后再从这个特征子集中选择最优特征用于结点的划分。

设数据集样本数为 n ，特征数为 m ，随机森林包含 T 棵树，则随机森林的流程如下：

- (1) 从包含 n 个样本的原始数据集中随机有放回地抽取 k (k 通常取 n) 个样本作为训练集。
- (2) 使用 (1) 中抽取的训练集训练一棵 CART 决策树，需要注意的是，这里训练决策树时无需剪枝。

(3) 重复步骤(1)和步骤(2) T 次, 得到 T 棵训练好的基决策树。由于各决策树训练过程独立, 因此可以并行化训练 T 棵树。

(4) 采用投票法(分类)或简单平均法(回归)将 T 棵基决策树的预测结果进行集成, 得到最终的预测结果。

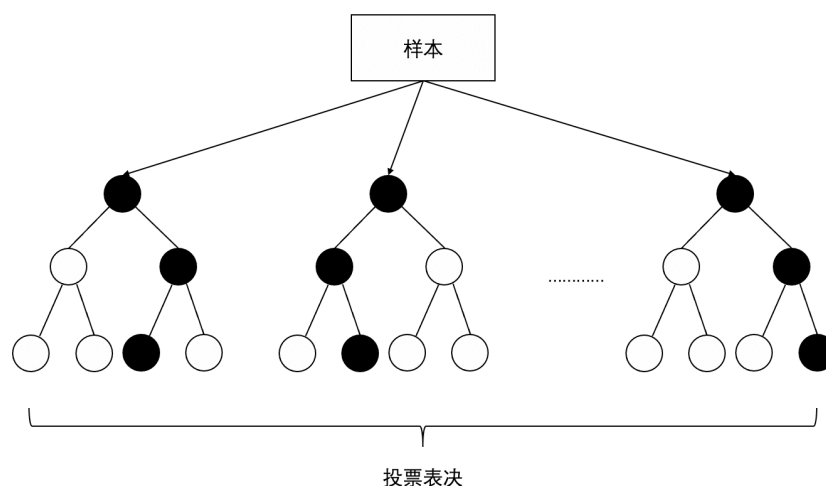


图 2.1 随机森林示意图

随机森林简单易实现, 由于各基决策树可以并行训练, 因此计算开销小。此外, 随机森林还具有以下特点:

(1) 随机森林的基学习器的多样性不仅来自样本随机性, 也来自属性随机性, 这就使得泛化性能可以通过基学习器之间差异度的增加而得到进一步提升。

(2) 随机森林的参数只有森林规模和结点分裂候选属性数, 超参数少, 简化了使用中的调参过程。

(3) 随机森林可以有效解决样本类别不平衡问题, 且特别适用于处理高维小样本数据。

(4) 随机森林可以处理离散变量和连续变量, 且不需要进行特征的标准化和归一化。

(5) 随机森林可以较好地处理缺失值, 且对异常值不敏感。

(6) 因为随机森林中每棵树的训练都只使用了部分样本, 因此在评估模型性能的时候可以直接使用各决策树未使用的样本进行袋外误差估计 (Out of Bag Estimation)。

(7) 随机森林在训练结束后不仅可以给出最终的预测结果，也可以给出各个特征的重要性得分，有助于后续进行特征选择。

正是因为随机森林具有上述诸多优点，它被广泛应用于很多场景的任务中，取得了良好的表现。随机森林被认为是集成学习中的代表性方法^[15]。

2.2 深度神经网络

深度学习 (Deep Learning) 是机器学习的一个分支，相比经典的统计学习模型，深度学习为模型引入了更大的“深度”(复杂性)，通过多个函数(层)来获得数据的抽象表示。深度神经网络 (Deep Neural Network) 是深度学习中最具代表性的方法。深度神经网络的一个强大优势是表征学习，即从原始数据中自动提取特征，并由较低层次的特征组成更高层次的特征^[16]。

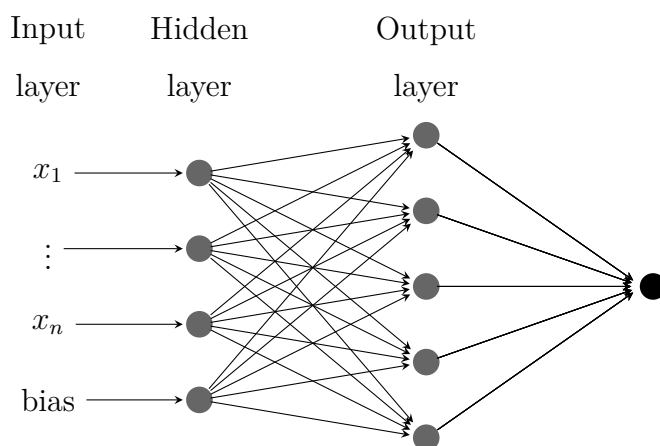


图 2.2 深度神经网络示意图

相比传统的统计学习模型，深度神经网络可以解决更复杂的问题，这一方面是因为深度神经网络模型具有很强的特征转换和提取能力，另一方面则是因为深度神经网络模型具有较高的模型复杂度。

深度神经网络也面临一定的问题，最常见的两类问题是过拟合和计算开销过大。过拟合往往是由数据量不足造成的，这也使得深度神经网络不太适用于数据较小的情况，为解决过拟合问题，可以在训练过程中采用权重递减 (L2 正则化)

或者稀疏 (L1 正则化) 等方法^[17]。计算开销大往往是由于网络结构复杂, 参数数量多造成的。为解决计算开销过大的问题, 可以使用小批量训练 (mini-batching) 的方法加速模型训练, 或设计适用于问题的特点网络结构来减少参数 (如 CNN)。

深度神经网络已被广泛应用于计算机视觉、语音识别、自然语言处理等领域, 并获取了极好的效果。

2.3 深度森林

如上所述, 随机森林具有训练快、实现简单、不易过拟合等优点, 但不具备深度神经网络强大的特征提取和转换能力; 深度神经网络具有可以自动提取特征、解决复杂问题的优点, 但不具备随机森林训练快、适用于小样本的特点。Zhou 和 Feng^[18] 则将两者的优势相结合, 提出了深度森林的方法。

作者首先分析了深度神经网络在很多场景表现优异的关键原因: (1) 对训练数据进行逐层处理; (2) 无需人工提取特征, 可以在模型内部进行特征转换 (即无需额外的特征工程); (3) 通常具有足够大的模型复杂度, 因此可以用于解决复杂问题。

接下来作者又分析了集成学习中通过增加基学习器的多样性来增强模型泛化能力的几个常见做法: (1) 对样本进行操作, 即随机对样本采样, 采用不同的样本子集作为训练集来训练各基分类器; (2) 对特征进行操作, 即随机对特征采样, 采用不同的特征子集来训练各基分类器; (3) 对各基学习器设置不同的超参数; (4) 对输出进行一定扰动。

受上述深度学习和集成学习的特点启发, 作者提出了 gcForest (multi-Grained Cascade Forest) 模型, 该方法将集成方法与深度神经网络的框架相结合, 既具备深度神经网络学习能力和表征能力强的优点, 又具备集成学习方法易于训练、有效防止过拟合的优点。gcForest 模型由级联森林 (Cascade Forest) 和多粒度扫描 (Multi-Grained Scanning) 两部分构成。

2.3.1 级联森林 (Cascade Forest)

深度学习中的表征学习依赖于对原始特征的逐层处理, 因此 gcForest 也采用了层级结构, 又称为级联森林结构 (Cascade Forest Structure)。所谓级联森林,

就是以类似神经网络的方式将多个随机森林分类器按层级结构组织在一起，每一层的随机森林都把前一层随机森林得到的输出作为输入，然后把处理得到的输出结果传递到下一层。

具体模型结构如图 2.3 所示。在图中的级联森林中，每一层都由 4 个随机森林分类器构成，这可以视作对随机森林这一集成方法的进一步集成。为保证森林的多样性，作者使用了 2 个普通随机森林和 2 个完全随机森林。其中普通随机森林在划分结点时随机选择 \sqrt{d} (d 为特征维数) 个特征作为特征子集，然后从中选择最优特征划分结点，而完全随机森林在划分结点时直接随机选择一个特征进行划分。

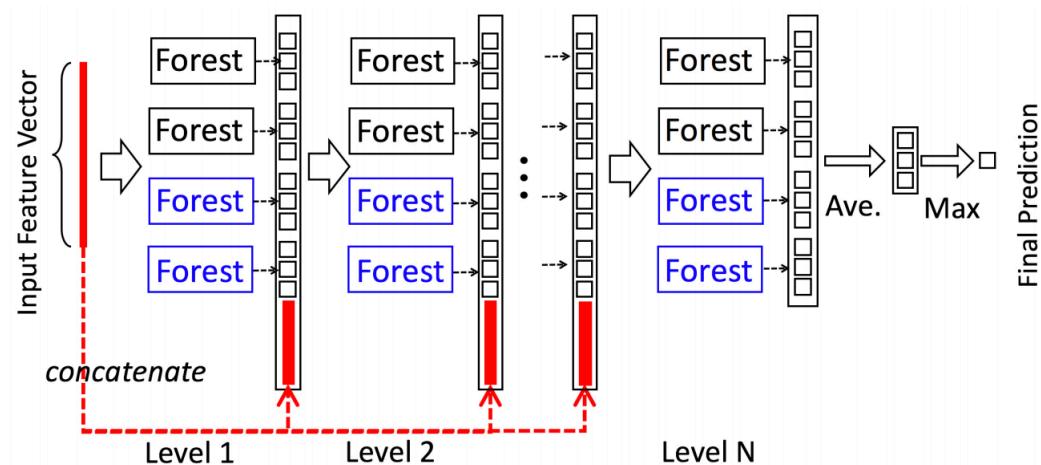


图 2.3 级联森林结构示意图^[18]

级联森林的训练过程如下：首先将样本划分成训练集和验证集，然后使用训练集去训练第一层中的随机森林，训练完成后，对输入模型的每个训练样本，找到该样本在各基决策树中所在的叶子结点，计算该叶节点上不同类别样本所占的百分比，作为该样本的类别分布向量，对同一随机森林中的所有基决策树给出的类别分布向量进行平均，得到该样本最终的类别分布向量。对每个样本，将各森林输出的类别分布向量拼接在一起，再与样本的原始特征拼接起来组成新的特征，作为下一层的输入。此外，我们需要记录模型在验证集上的分类表现，并与加入下一层后模型在验证集上的分类表现进行比较，若加入下一层后的模型表现变差，则去掉下一层，并停止训练，否则加入下一层，重复上述训练过程。

级联森林训练结束后，我们将测试集的样本输入到模型中去进行预测。与训练过程类似，各层训练好的随机森林输出样本的类别向量特征并与原始特征拼接传递到下一层，到达最后一层时，对各随机森林输出的类别分布向量求简单平均就可以得到样本属于各类的概率，最大概率对应的类别就是级联森林模型的最终预测结果。

可以看到，在级联森林的训练和预测过程中，类别分布向量起到了十分重要的作用。因为各决策树从根结点到叶子结点的路径可以视作一种特征转换，因此可以认为样本落入的叶子结点包含了样本的高阶特征信息，从而类别分布向量可以视为转换后得到的新特征。将类别分布向量与原始特征进行拼接并传递到下一层的过程可以视作不断提取高阶特征的过程（与神经网络类似）。类别分布向量的计算过程如图 2.4 所示：

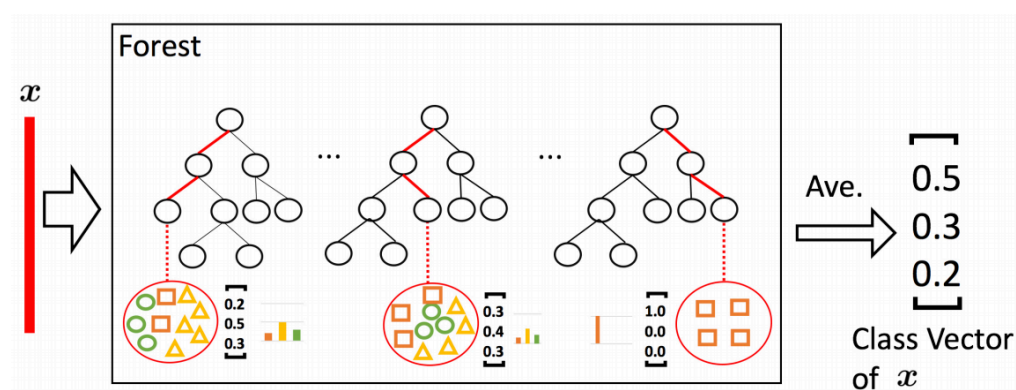


图 2.4 类别分布向量计算过程示意图^[18]

作者指出，这里采用类别分布向量作为各层新提取的每个增广特征是一种非常简单的做法，如此少量的增广特征所提供的增广信息非常有限，当原始特征向量维度很高的时候，增广特征很可能会被原始特征淹没。但作者通过实验表明，即使采用这样一种简单的特征增广方式，也已经可以使得级联森林模型具有较好的表现。作者认为更多形式的增广特征可能会提取出更多有用的特征信息，从而产生更好的效果。

以上就是模型的框架和训练流程，可以看到整个流程类似于一个可以跨层传递信息的前馈神经网络。但神经网络是端到端进行训练的，而级联森林却需要自

发的停止层的生长。为避免过拟合，增加一个新的层后，要在验证集上对整个级联森林的性能进行估计，如果没有显著的性能增益，则训练过程终止。与大多数深度神经网络的模型层数固定不同，级联森林的层数是自动确定的。gcForest 通过在适当的时候终止训练来自适应地确定其模型复杂度，这使得它可以适用于不同规模的训练数据，而不局限于大规模的训练数据。

2.3.2 多粒度扫描 (Multi-Grained Scanning)

深度神经网络具有很强的特征提取和特征转换能力，但处理高维数据时，可能会面临参数数量多、计算开销大的问题，因此针对具有特殊结构的高维数据，可以设计一些特殊的网络结构来更好地提取特征，比如卷积神经网络 (CNN) 被广泛应用于图像数据的处理^[19]，循环神经网络 (RNN) 被广泛应用于语音信号等序列数据的处理^[20]。这些人为设计的特殊网络结构能够以更小的计算开销提取出更多我们关注的特征信息，如卷积神经网络可以关注图像不同位置的局部特征，然后逐层抽象出整张图像的高阶特征，循环神经网络则可以关注序列数据的上下文联系，提取出对全局理解更有帮助的高阶特征。

对于 gcForest 模型而言，也需要一种方法来对高维的、具有一定结构特点的数据进行处理，然后再将处理后的数据输入到级联森林中去。这是因为对于图片或语音信号这样具有特殊结构的数据，直接使用级联森林模型难以很好的提取和整合局部特征。

为了解决上述问题，作者提出了多粒度扫描 (Multi-Grained Scanning) 的方法，该方法受卷积神经网络的启发，使用类似卷积的窗口在具有特殊结构的数据上进行滑动，按照窗口大小将原始数据每个样本都拆分成多个子样本，子样本标签与原样本一致。使用所有的子样本训练随机森林分类器，并仿照级联森林中的方式，通过将森林输出的类别分布向量拼接在一起来实现对原始数据局部特征的提取和整合。多粒度扫描的具体流程如下：

(1) 对每个 p 维的样本，首先通过一个长度为 k 的采样窗口进行滑动采样，得到 $s = (p - k) + 1$ 个特征维数为 k 的子样本。

(2) 将每个子样本打上和对应的原样本相同的标签，并将这些子样本用于完全随机森林和普通随机森林的训练。

(3) 每个训练好的森林对每个子样本都输出一个长度为 c 的类别分布向量 (c 为类别数量)，对每个 p 维的样本，每个森林都输出一个 $s * c$ 的表征向量，最后把多个森林的输出拼接在一起作为后续级联森林的输入。

需要注意的是，上述步骤 (2) 将每个子样本打上和对应的原样本相同的标签的做法看起来并不合理，因为原始特征中未必每个局部都包含我们关注的与类别相关的信息，比如一张汽车图片中未必每个局部都有汽车相关的特征，作者认为这里可以将其视作对输出进行的扰动，与 Flipping Output^[21] 方法类似，某种程度上这样的做法可以增加集成的多样性。而增加集成多样性有利于提升集成分类器的整体表现。

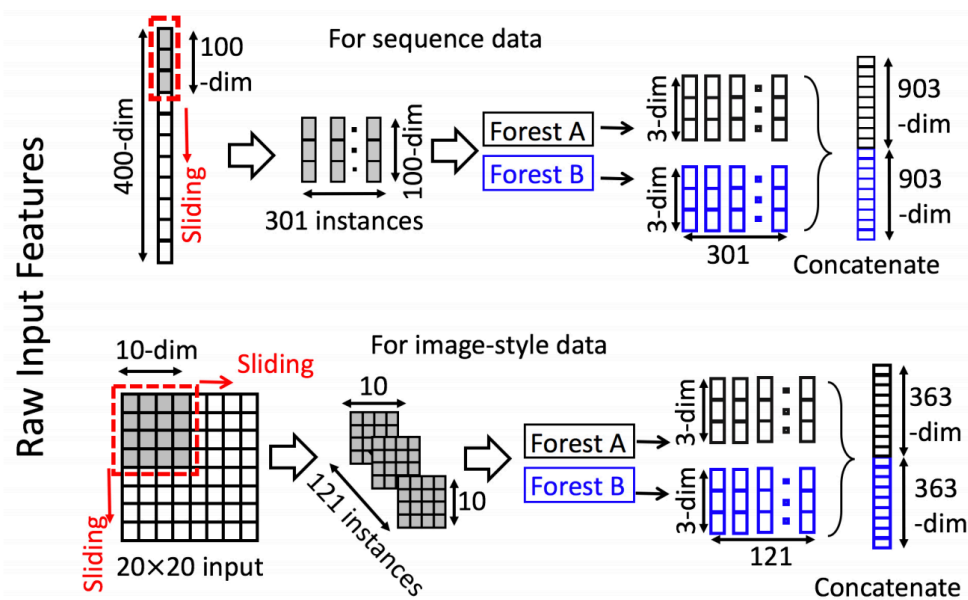


图 2.5 多粒度扫描流程示意图^[18]

图 2.5 展示了两个多粒度扫描的示例。在第一个示例中，我们处理的数据是 400 维的序列数据，这里采用的采样窗口长度为 100，步长为 1。扫描后得到 301 个 100 维的子样本，将这些样本输入两个森林分类器（包括一个普通随机森林和一个完全随机森林）进行训练，每个随机森林将输出的 301 个 3 维的类别分布向量拼接在一起得到一个 903 维的特征向量，将两个随机森林的输出拼接在一起得到 1806 维的特征向量，作为之后级联森林的输入。第二个示例中，我们处理的

数据是 20×20 的图像数据，这里采用的采样窗口大小为 10×10 ，步长为 1。扫描后得到 $11 \times 11 = 121$ 个 10×10 维的子样本，将这些子样本输入两个随机森林分类器进行训练，每个随机森林将输出的 121 个 3 维的类别分布向量拼接在一起得到一个 363 维的特征向量，然后将两个随机森林的输出拼接在一起得到 726 维的特征向量，作为之后级联森林的输入。多粒度扫描可以视作对数据的预处理，目的是提取并整合局部特征用于后续级联森林的训练。

需要注意的是，当变换后的特征向量太长时，我们可以采取一些措施来减少特征维数，如：（1）对滑动窗口扫描生成的实例进行子采样。该采样过程与随机子空间方法^[22]类似，可用于增加集成多样性；（2）选取更大的窗口滑动步长，这样在减少特征维数的同时也可以降低计算开销。

此外，与卷积神经网络可以采用不同大小卷积核一样，多粒度扫描中的窗口大小也可以设置多个，从而获得不同大小的感受野。在实际应用中，我们可以将采用不同大小的窗口产生的特征分别作为级联森林不同层的输入，下一节中给出了一个具体的示例。

2.3.3 深度森林 (gcForest)

gcForest 模型由上述的多粒度扫描和级联森林两部分组成，完整流程图如图 2.6 所示。

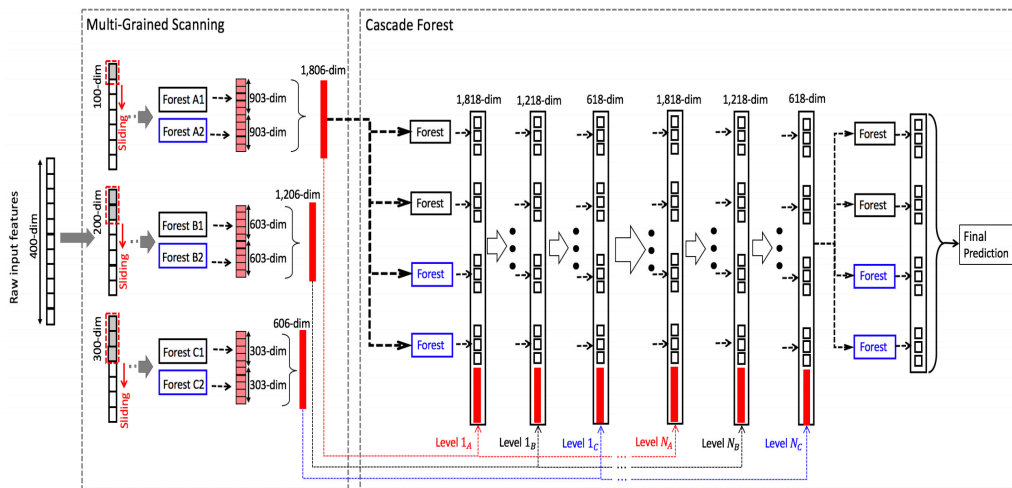


图 2.6 深度森林完整流程图^[18]

gcForest 的完整流程如下：给定一组训练样本作为输入，首先通过多粒度扫描过程来获得各个样本对应的转换后特征表示，然后将其作为级联森林的输入（若进行了多次多粒度扫描得到了多个特征表示，则可以作为级联森林不同层的输入，如图 2.6 所示）。当然对于没有特殊结构的数据可以省略多粒度扫描过程。使用级联森林进行训练，级联森林在训练过程中不断生长直至达到停止条件（例如训练误差不再减少或训练误差减少值低于某个阈值）。然后将测试集样本输入到训练好的 gcForest 模型中，经过各层的特征转换直至最后一层，对模型最后一层输出的多个类别分布向量求简单平均，其中最大概率对应的类别即为整个 gcForest 模型最终的预测结果。

表 2.1 gcForest 模型中的超参数

多粒度扫描	级联结构
多粒度扫描的森林数	级联的每层的森林数
每个森林的决策树数	每个森林的决策树数
树的停止生长规则	树停止生长的规则
滑动窗口的数量和大小	

总的来说，gcForest 模型具有以下几个优点：

(1) 计算开销小，运行速度快。相比深度神经网络，gcForest 计算开销通常较小。随机森林可以并行化训练，因此 gcForest 运算速度一般较快。

(2) 模型效果好。论文^[18]通过实验对比了 gcForest 和经典机器学习方法以及深度神经网络在多个分类任务中的表现，验证了 gcForest 具有很好的效果。

(3) 超参数少，模型对超参数不敏感。不同于深度学习的网络结构参数，gcForest 的超参数可以不用调参，或仅简单地调整。gcForest 的超参数见表 2.1。

(4) 级联森林的生成依赖于增加一层后模型性能的变化，避免了过拟合。模型复杂度可以自动适应不同规模的数据集。

(5) 相比深度神经网络，gcForest 模型更容易进行理论分析。

目前，gcForest 模型已被应用到不同领域的很多分类任务中，并取得了不错的表现^{[23][24]}。

2.4 局部二值模式

局部二值模式 (Local Binary Pattern, 简称 LBP) 最初是 Ojala 等人在 2002 年提出的一种图像纹理描述方法^[25]。简单来说, LBP 就是通过将每个像素的相邻像素与该像素值进行比较, 并将其作为二进制编码来表示局部纹理。由于 LBP 在很多场景的应用中都有良好的表现, 尤其是在计算机视觉和图像处理领域, 因此得到了研究人员的广泛关注。LBP 通过对各中心像素与相邻像素进行简单的计算来刻画图像中存在的纹理模式。计算的结果通常被汇总成直方图形式, 作为刻画图像纹理的特征使用。由于二维 LBP 在获取图像基本纹理特征方面计算简单且效果好, 因此 LBP 的一维版本 (简称 1DLBP) 近年来也被用于一维信号的特征提取^[26]。

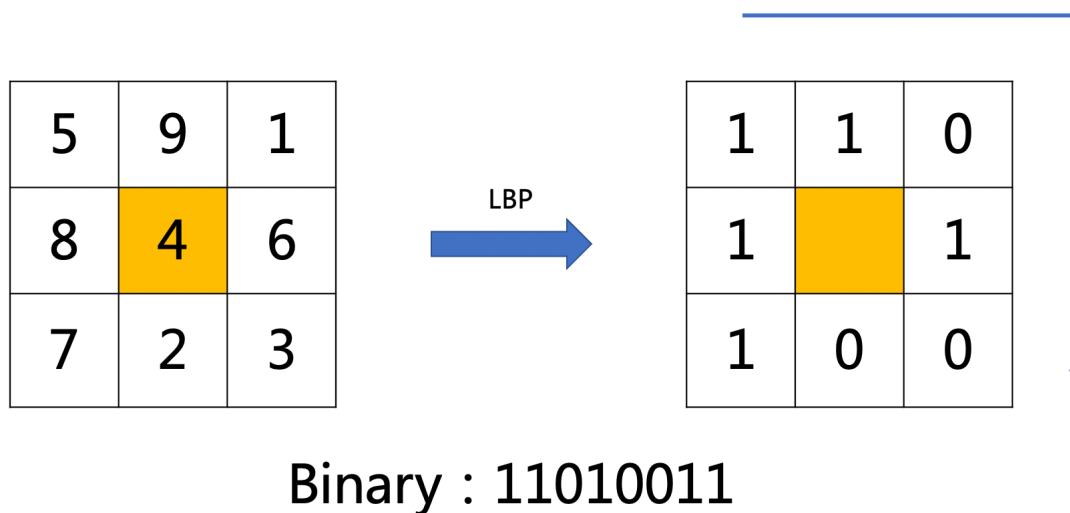


图 2.7 2DLBP 示意图

2DLBP 计算特征的过程示例如图 2.7 所示。使用 2DLBP 从数据中提取特征向量的步骤如下:

(1) 将图片划分成多个窗口 (如 16×16 像素的窗口)。

(2) 对窗口中的每个像素, 将其与 8 个相邻的像素值进行比较。若中心像素大于某个邻域的, 设置为 0; 否则设置为 1。这就获得了一个 8 位的二进制数 (通常会转换为 0~255 的十进制数), 作为该位置的特征。

(3) 对每个窗口的特征计数得到直方图，并将所有窗口的特征直方图组合在一起，若样本维度不统一则需要对直方图进行归一化，这样就得到了当前图片的特征向量（在本例中特征向量维度为 256）。

1DLBP 的计算过程与 2DLBP 算法类似。实际上，对于一维的信号数据，我们将中心振幅值与左右各 4 个共 8 个（也可以是其它数值，通常默认值为 8，与二维的情形一致）相邻振幅值进行比较，若某个相邻振幅值小于中心振幅值，则该位置记为 0；若某个相邻振幅值大于中心振幅值，则该位置记为 1。将上述 0、1 组成的二进制数转化为 0~255 的十进制数作为该位置的特征，遍历整个信号数据后我们对所有特征值进行计数就可以得到一个直方图，最终我们把得到的直方图（256 维）作为当前信号数据的特征向量，就完成了特征提取过程。1DLBP 提取特征的过程如图 2.8 所示：

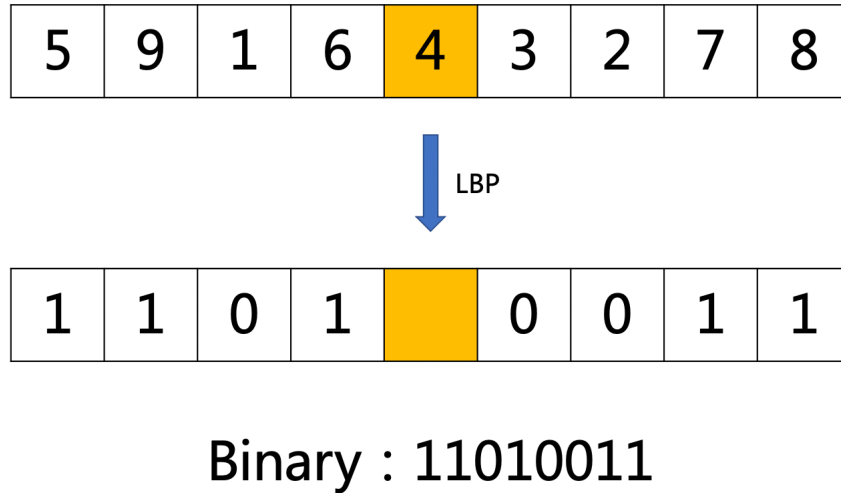


图 2.8 1DLBP 示意图

这里我们也可以给出 1DLBP 运算的数学定义：

$$1DLBP_P^{\text{All}} = \sum_{p=1}^P \zeta(S_p - S_c) 2^{p-1} \quad (2.1)$$

$$\zeta(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

在使用 1DLBP 处理一维的信号数据时，我们就可以使用一个定长的窗口沿数据滑动，对每个窗口内都进行式 (2.1) 的运算得到特征值，最后将各特征值计数得到直方图作为从原始数据中提取出的特征向量。1DLBP 算法在本文的 sEMG 手势分类问题中将被用作时域特征的提取。

2.5 本章小结

本章首先对随机森林和深度学习的原理和特点进行了简要介绍，然后介绍了 Zhou 和 Feng 提出的结合随机森林和深度学习优点的模型——深度森林，包括深度森林的思想、流程以及优势。此外，在本章末尾，给出了局部二值模式 (LBP) 算法的介绍，这一算法之后将用于表面肌电信号时域特征的提取。

第3章 改进的深度森林算法

Zhou 和 Feng^[18] 提出的深度森林 (gcForest) 模型将随机森林构造成层级结构, 在很多分类任务中达到了优于传统机器学习算法和深度神经网络的表现。根据上一章中的介绍, 我们知道 gcForest 模型尤其适合处理高维小样本数据, 而基于 sEMG 进行手势分类就是一个高维小样本数据上的多分类问题, 因此使用 gcForest 来解决该问题是合适的。为了使 gcForest 模型能够更加充分地从高维小样本数据中提取有用信息来构造高阶特征, 进而提升模型表现, 本文对 gcForest 模型进行了一些改进。需要注意的是, 虽然我们改进 gcForest 模型的出发点是使其更加适用于处理高维小样本数据上的多分类问题, 但改进后的方法在其它类型的问题中也可以取得较好的表现, 下一章中我们通过实验来验证模型的效果。

3.1 带类别权重的随机森林

在 gcForest 模型的级联森林组件中, 每一层提取的特征由各样本通过该层的随机森林分类器后输出的多个类别分布向量拼接构成, 并且除了第一层之外, 每一层的输入都是由原始特征和上一层新提取出的特征拼接而成。虽然采用这样的做法已经可以在很多实验中取得较好的表现, 但该方法也存在一定问题:

(1) 各层中随机森林的多样性依赖于人为指定。如在 gcForest 中, 每层的随机森林包括 2 个随机森林和 2 个完全随机森林, 这是由作者设置的, 这样的做法为集成提供的多样性有限, 可能无法提取出充足的特征信息。尤其对于小样本的数据来说, 训练具有相同超参数的两个随机森林提取出的特征信息可能较为相似, 从而使得级联森林提取特征的能力较弱。

(2) 当原始特征维数较高时, 每层新提取出的特征维数比原始特征维数小很多, 这会使得下一层各随机森林的基决策树在进行训练时, 很多结点划分所使用的特征子集都几乎不含前一层新提取出的特征, 导致特征淹没问题。特征淹没又会使得级联森林提前停止生长, 从而导致模型结果较差。这一问题在处理特征维数很高的数据时尤为明显。

为了解决上述问题，本文提出了一种改进方法——在级联森林各层中使用带类别权重的随机森林。具体思路如下：在级联森林的每一层中，都训练 c 个随机森林（ c 为样本类别数），训练各随机森林时分别对训练样本赋予不同的类别权重。具体来说，当训练第 k 个森林时，我们赋予第 k 类的样本一个较大的权重，使得该随机森林在训练时重点关注第 k 类的样本和其它类别样本之间的差异，所谓重点关注，其实就是在训练各基决策树时，从根节点开始的前几次结点划分会优先考虑将第 k 类的样本和其它类别的样本区分开，因此最终随机森林提取出的特征也更能反映第 k 类的样本和其它类别样本之间的差异。通过这样的方式训练出的 c 个随机森林在提取特征方面更加互补，因此对特征信息的提取也更加充分。这里给出一个简单的例子来说明这一点：在手写数字分类的问题中，给予数字“8”较大权重的随机森林会重点关注数字“8”与其它数字的差异，因此可能会学到“8”区别于其它数字的结构特征“×”。同理，给予其它数字较大权重进行训练时，森林也更可能学习到该数字独有的特征。图 3.1 给出了一个使用带类别权重的随机森林解决三分类问题的示例：

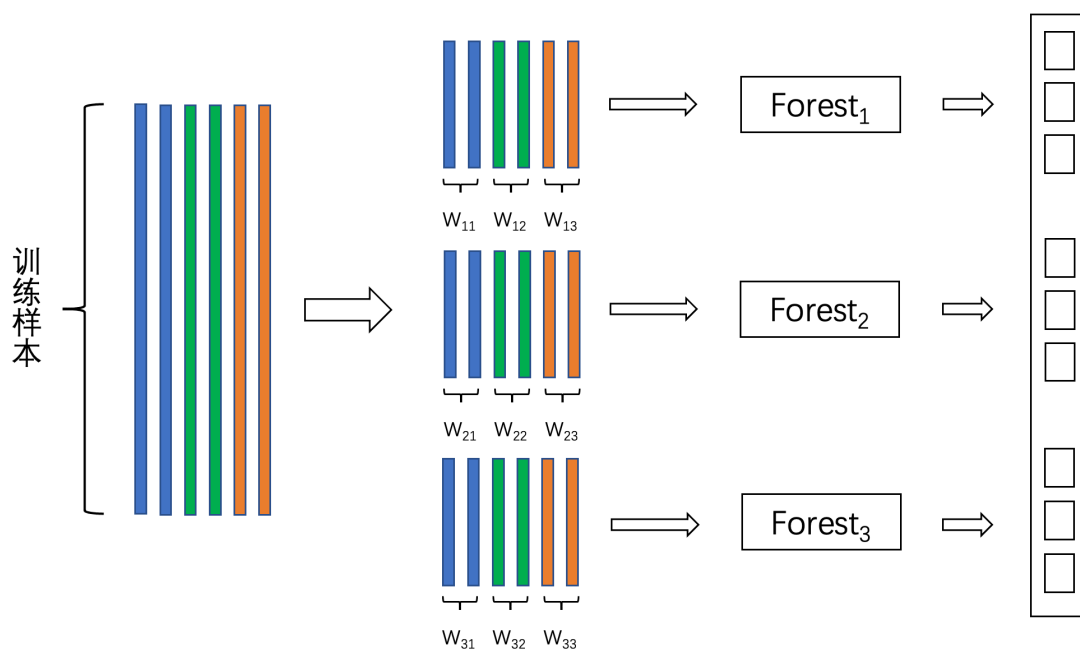


图 3.1 单层中带类别权重的森林示意图

可以看到, 对于图 3.1 中的三分类问题, 每层森林的训练涉及 9 个权重参数, 对于 c 分类问题, 则每层森林的训练涉及 c^2 个权重参数。为简单起见, 我们采用一种朴素的赋权方法:

$$w_{ij} = \begin{cases} \alpha(\alpha > 1), & \text{if } i = j \\ 1, & \text{if } i \neq j \end{cases} \quad (3.1)$$

这样一来, 相比 gcForest 模型, 我们多了一个权重超参数 α , 同时少了一个每层森林数的超参数 (因为我们的方法中每层的森林数由类别数确定), 因此超参数总数不变。

使用带类别权重的随机森林构成的级联森林相比 gcForest 中原始的级联森林具有以下优点:

(1) 每层中随机森林分类器的多样性不再只依赖于人为设定, 而是同时依赖于各随机森林输入的训练样本的多样性 (各森林输入样本的权重不同), 这将有助于级联森林模型自适应于处理的数据集。

(2) 改进后的级联森林每一层新提取的特征数目为 c^2 (c 为类别数), 当类别较多 ($c > 4$) 时, 改进后的级联森林相比 gcForest 中原始的级联森林每层可以提取更多维度的特征, 一定程度上缓解了特征淹没问题。

使用带类别权重的随机森林的缺点也很明显, 那就是需要消耗更多的计算资源, 尤其是样本类别较多的时候。这里我们可以通过采取一些措施来改善这个问题, 比如减少各随机森林中基决策树的个数, 这样做是因为当我们重点关注某一类的样本和其他类别样本差异的时候, 实际上我们可以将其大致视为一个二分类问题, 因为各决策树的前几次结点划分一定是优先将该类样本与其它类样本分开, 当一个结点不再包含或只包含很少的该类样本后, 结点划分才会考虑其他类别样本之间的差异。而二分类相比原问题中的多分类复杂度更低, 因此需要的基决策树数量也就更少。此外我们也可以通过限制各随机森林的最大树高等措施来进一步降低计算开销。

总而言之, 通过使用带类别权重的随机森林来构造级联森林, 可以进一步增加集成多样性、更加充分地提取特征、一定程度上缓解特征淹没问题。改进后的级联森林更加适合解决基于 sEMG 的手势分类问题。

3.2 序列化训练的随机森林

如上所述,通过使用带类别权重的随机森林来构造级联森林,可以让我们重点关注每个单一类别的样本和其他类别样本之间的差异,从而提取到各类别样本特有的某些特征信息,使得提取的特征更加充分、更加互补。但需要注意的是,gcForest 模型中使用不带权重的样本训练的随机森林依然有其保留的必要性,这是因为使用不带权重的样本训练的随机森林关注所有类别之间的差异大小,因此训练出的随机森林与使用带类别权重的样本训练的随机森林之间本身就具有一定的差异性。以手写数字识别问题举例,在使用不带权重的样本训练的随机森林中,“0”、“6”、“8”、“9”几个数字可能由于都具有“0”这一结构被划入同一个结点中,然后再进行进一步的划分,而使用“8”类别权重较大的样本训练的随机森林在时可能会在一开始就采用“×”结构先将“8”类别的样本与其它数字分开。因此保留 gcForest 中原有的随机森林有助于进一步增加集成的多样性。

接下来需要考虑的问题是:我们是否直接保留 gcForest 各层中所有的随机森林呢?关于这个问题我们上一节已经提到,当几个随机森林超参数一致时,使用相同的样本训练得出的各随机森林之间的差异性较为有限。因此我们希望通过某种方式使得各层中的随机森林在提取特征方面具有较大差异性。

受到 Boosting 方法的启发,我们可以序列化训练随机森林使得各随机森林提取的特征信息更加互补,实际上,Facebook 于 2014 年提出的经典的 GBDT (Gradient Boosting Decision Trees) +LR(Logistic Regression) 的 CTR 模型^[27]就提供了很好的特征工程思路。仿照 Boosting 的做法,我们可以先训练一个随机森林,再训练第二个随机森林并令其关注第一个随机森林误分类的样本,但是通过实验检验我们发现这样的做法效果并不是很好,可能的理由是随机森林不是弱学习器,因此并不能像决策树一样简单地进行提升学习。

需要明确的是,我们这里训练多个随机森林并不是为了将它们的预测结果集成来对样本进行预测,而是希望把随机森林作为特征提取或特征转换模型。从这个角度来说,我们无需采用提升的方法来处理随机森林,但可以借鉴其序列化训练森林的处理过程。本文提出的方法流程如下:

(1) 先使用原始样本作为输入训练第一个随机森林,由训练好的随机森林得到特征重要性得分。

(2) 去掉特征重要性得分最高的几个特征（同时也可以去掉重要性得分最低的几个特征），并使用剩下的特征来训练第二个随机森林。

(3) 将样本输入两个训练好的随机森林，输出的类别概率向量拼接在一起即为提取出的特征。

序列化训练两个随机森林的流程如图 3.2 所示。

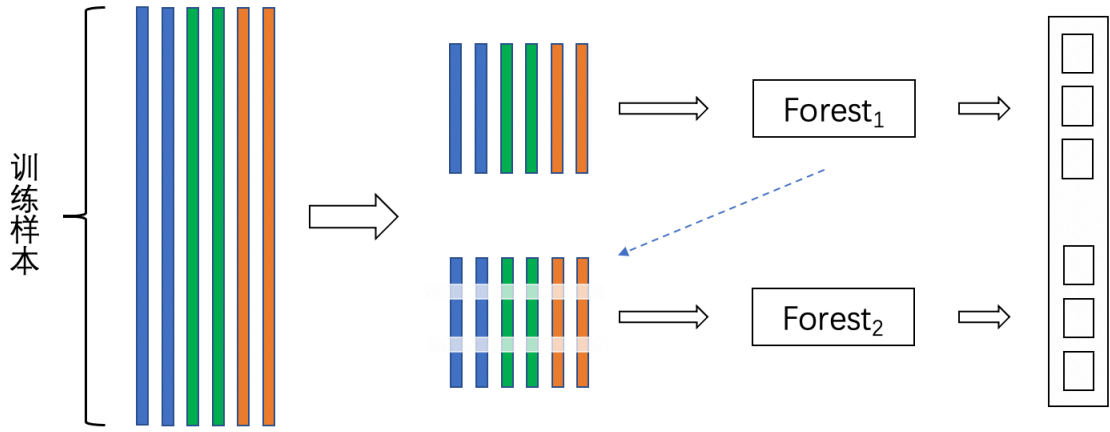


图 3.2 单层中序列化训练的随机森林示意图

上面我们给出了算法的流程，接下来我们给出更为详细的形式化表述。设样本特征集为 $f = \{f_1, f_2, \dots, f_m\}$ ， m 为特征维数。训练第一个随机森林使用的样本特征为 $F_1 = f$ 。第一个随机森林训练完成后，可以得到特征特征集 f 的特征重要性得分，记为 $score = \{score_1, score_2, \dots, score_m\}$ 。对 $score$ 降序排列得到按重要性排名的特征列表 $f' = \{f'_1, f'_2, \dots, f'_m\}$ 。然后我们可以选择去掉排名前 β 的特征（ β 可以取各决策树结点划分时选取的特征子集的大小，通常为 \sqrt{m} ），并使用 $F_2 = \{f'_{\beta+1}, f'_{\beta+2}, \dots, f'_m\}$ 来训练第二个随机森林。或者我们也可以去掉排名前 $\beta/2$ 和后 $\beta/2$ 的特征，即使用 $F_2 = \{f'_{\frac{\beta}{2}+1}, f'_{\frac{\beta}{2}+2}, \dots, f'_{m-\frac{\beta}{2}}\}$ 来训练第二个随机森林。两个随机森林训练完成后，将各自的输出拼接到一起就得到了从输入中提取到的特征。

这里我们采用的方案比较简单，即直接去掉特征重要性最高的部分特征（同时也可以去掉重要性最低的部分特征），但实际上我们也可以选择从最重要的特征中随机去掉部分特征（同时也可以从最不重要的特征中随机去掉部分特征）。

经过简单的实验验证,我们发现在使用超参数相同的2个随机森林时,相比直接把原始特征作为输入且并行化训练的随机森林,按上述步骤序列化训练的随机森林表现要更好。在某些任务中,序列化训练2个随机森林与并行化训练4个随机森林的表现相当,有时甚至前者比后者的表现要更好。

那么这种序列化训练森林的思路为什么能够产生较好的效果呢?其实某种程度上我们采用的方法和 Hinton 于 2012 年提出的 Dropout 方法^[28]有类似之处。Dropout 是神经网络进行前向传播的时候,以一定的概率 p 停止神经元的激活过程,每次都只训练网络中的部分权重。实际上,这个过程可以视作我们每次只用部分特征去构造更高阶的特征以使得预测结果与真实标签接近,这样做可以有效避免过拟合。直觉上,如果我们使用所有特征进行训练,则可能有部分重要特征的权重会得到充分的训练,而其它非重要特征的权重可能得不到充分的训练,但如果我们把某些重要特征对应的神经元“灭活”,则非重要特征的权重可以得到更好的训练,从而可能提取出新的有用的特征信息。我们这里采用的方法也是类似的思路,但因为随机森林没有办法像神经网络那样进行多轮次的训练,因此我们序列化地训练随机森林,即对第一个随机森林中得到的重要特征进行“灭活”后来训练第二个随机森林。这样可以使得两个随机森林差异性较大,提取出的特征较为互补。

上面我们解释了为什么对第一个随机森林中得到的重要特征进行“灭活”后再训练第二个随机森林可以取得较好的效果,下面我们简单说下去掉重要性得分最低的部分特征的原因。实际上当我们去掉重要性得分最高的部分特征后,剩下的特征中非重要特征占比会增加,这就导致随机森林的基决策树在进行结点划分时选取的特征子集会包含较多的非重要特征,从而影响模型效果,设原特征中重要特征数为 a ,非重要特征数为 b ,现要去掉 c 个特征,则显然 $\frac{a-c}{a+b-c} > \frac{a-\frac{c}{2}}{a+b-\frac{c}{2}}$,因此去掉重要性得分最低的部分特征可以使得第二个随机森林在与第一个随机森林保持较大差异性的同时具有较好的表现。举例来说,在手写数字识别的问题中,第一个随机森林得到的重要性得分最低的像素特征都位于图像四角的空白区域,去掉这部分特征会使得第二个随机森林在训练时可以关注更有用且与第一个随机森林差异化的特征转换。

3.3 层级间特征选择机制

在 gcForest 的级联森林中，每一层的特征向量都由原始特征向量和上一层新提取出的特征向量拼接而成，这样做的好处在于每一层中都完整地保留了原始特征信息，从而每一层提取向量时都相当于从原始特征中提取有用的信息与上一层提取的高阶特征组合成更高阶的特征，然后传递到下一层。

将该过程与传统的前馈神经网络相比，我们可以发现 gcForest 的做法更加谨慎，前馈神经网络每一层只保留上一层提取出的特征，而 gcForest 则始终保留原始特征，这样做主要是因为：

(1) 森林作为一种特征提取的工具没有神经网络那么强大，神经网络实现了端到端的训练，能够以最终的目标标签为导向不断修正每一层提取出的特征，而 gcForest 每一层中提取的特征都不会受到之后层的影响，每一层训练完毕后也不会再进行修正，因此保留原始信息可以给各层修正之前提取的特征的机会。

(2) 森林提取出的特征维数较低，若不加入原始特征，则提取特征的过程会损失大量的信息。

虽然 gcForest 采用始终保留原始信息的做法有上述的必要性，但这样的做法也会导致一些问题：

(1) 由随机森林提取出特征维数较低导致的特征淹没问题，可能使得提取特征时难以充分利用上一层中提取到的高阶特征信息。因为随机森林在训练基决策树时会先随机选取特征子集，而特征子集不包含上一层提取的特征的概率较大时，就会产生上述问题。

(2) 当原始特征存在较多无关或冗余特征的时候，保留原始特征的做法会导致训练受阻，可能使得级联森林过早停止生长，使得模型提取不到高质量的高阶特征。

为了解决上述问题，且在保留原始信息与充分利用各层提取的高阶特征之间找到平衡，本文提出了层级间的特征选择机制，即级联森林中每一层的特征向量传递到下一层时都先进行特征选择，将选择后的特征向量与下一层新提取出的特征向量进行拼接得到下一层的特征向量。这样的做法使得每一层都可以剔除掉一些重要性较低的特征，并保留一些之前层提取出的较为重要的高阶特征。该做法具有以下优点：

(1) 可以使得各层中提取出的高阶特征传递到后几层，而不仅仅传递到后一层，这有助于更高阶特征的提取。

(2) 在每一层中都剔除一些与任务无关的冗余特征或重要性较低的特征，可以加快模型的收敛速度。

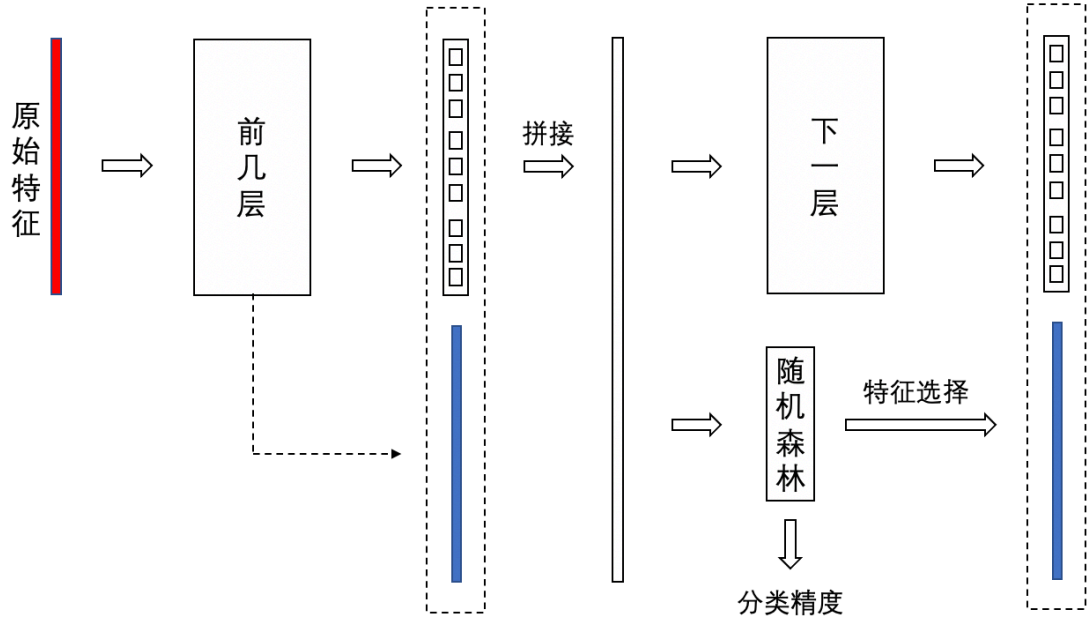


图 3.3 层级间特征选择机制示意图

层级间特征选择的流程如图 3.3 所示。这里我们采用随机森林用于特征选择，这样可以节约计算资源。具体来说，在 gcForest 的级联森林结构中，各随机森林分类器输出的类向量本身就是提取出的特征，因此随机森林既用于提取特征，同时又用于计算分类精度来评估模型表现。而在本文提出的修改后的结构中，带类别权重的随机森林都更加侧重某些样本的学习，而序列化训练的随机森林中第二个森林是为了提取与第一个森林更加互补的特征，因此两者得到的类向量都不适合直接用于评估模型表现。因此我们把上述随机森林视作专门提取特征的随机森林，在此基础上引入新的随机森林，专门用于计算分类精度评估模型表现。这些评估模型表现的随机森林训练好后，直接可以给出特征重要性得分，帮助我们进行特征选择。

为了使得结果更加准确，我们通常采用两个或两个以上的随机森林用于模型评估以及特征提取。假设一共使用了 n 个森林，则对于每个特征 $f_i (i = 1, 2, \dots, M)$ ，都能得到 n 个重要性得分 $\lambda_{ij} (j = 1, 2, \dots, n)$ ，这里 M 为特征维数。我们通过计算特征重要性得分的平均值来得到最终的重要性得分：

$$s_i = \frac{1}{n} \sum_{j=1}^M \lambda_{ij} \quad (3.2)$$

最后将特征按照重要性得分 $s_i (i = 1, 2, \dots, M)$ 进行降序排列，取排名前 m 的特征即可。在本文的方法中， m 取原始特征的维数。

3.4 改进方法的整体框架

在前三节中，为了使得 gcForest 模型更加适用于高维小样本数据上的多分类问题，我们基于 gcForest 模型提出了 3 点改进，并简单说明了改进的原因。现将这三点改进进行汇总，得到如图 3.4 所示的完整流程图：

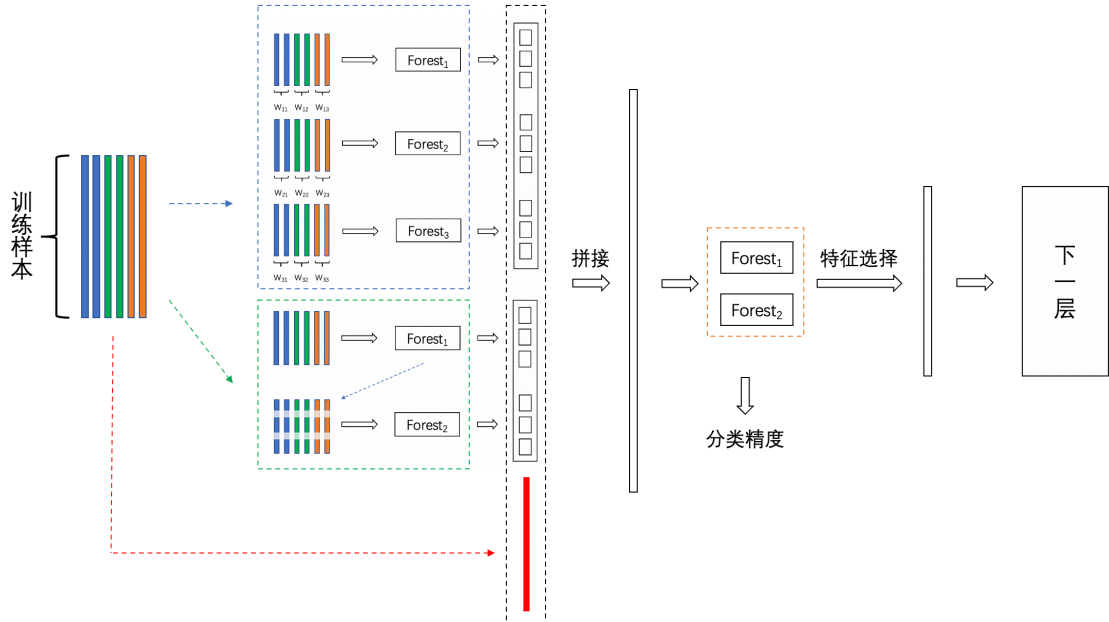


图 3.4 改进方法的整体框架

假设我们的训练集有 n 个样本，特征维数为 m ，现在我们要解决一个 c 分类问题。则整个流程如下：

(1) 将训练样本 v_0 输入模型，训练 c 个随机森林，每个森林重点关注一类样本，提取出 c^2 维的特征向量 v_1 。

(2) 将训练样本 v_0 输入模型，序列化训练 2 个随机森林，提取出 $2c$ 维的特征向量 v_2 。

(3) 将 v_0 、 v_1 、 v_2 拼接得到 $m + c^2 + 2c$ 维的特征向量 v_3 ，使用 v_3 训练两个随机森林并计算特征重要度排序。选择特征重要度最高的 m 个特征作为下一层的输入。

(4) 将验证集的样本输入到训练好的模型中，重复 (1)~(3)，唯一差别在于最后两个随机森林不用于特征选择而用于计算分类精度。将验证集在当前层的精度与前一层的精度进行比较，若精度下降则停止增长。

(5) 将测试集输入到训练好的模型中，将最后一层森林输出的类别分布向量做简单平均，得出模型对样本属于各类别概率的预测结果。

改进后模型的超参数如表 3.1 所示：

表 3.1 改进后模型中的超参数

多粒度扫描	级联结构
多粒度扫描的森林数	带类别权重森林中的权重参数 α
滑动窗口的数量和大小	序列化训练森林中的参数 β
每个森林的决策树数	每个森林的决策树数
树的停止生长规则	树停止生长的规则

以上就是改进方法的整体框架以及算法流程。相比 gcForest 方法，本文提出的改进方法具有以下优点：

(1) 同一层训练的各个用于提取特征的随机森林具有更大的差异性，且该差异性不依赖人为设定，而是依赖于数据，从而使得集成效果更好，可以更加充分地提取特征信息。具体来说，差异性来自于以下几个方面：a、同一层中的随机森林可分为使用带类别权重样本训练的随机森林和使用原始样本训练的随机森林，

两者侧重不同，提取出的特征具有较大差异性；b、使用带类别权重的样本训练的各随机森林之间具有较大差异性，这是因为每个随机森林都重点关注某一类别和其它类别之间的差异；c、使用原始样本训练的随机森林之间具有较大差异性，这是因为随机森林是序列化训练的，第二个随机森林关注第一个随机森林中不太重要的特征。

(2) 各层之间的特征传递不是简单的将前一层提取的特征和原始特征进行拼接作为下一层的特征，而是先进行特征选择再传递给下一层，这一措施有利于高阶特征的提取并加速模型的收敛。

(3) 对于多分类问题，改进方法可以一定程度上缓解特征淹没问题。

3.5 本章小结

本章主要阐述了原深度森林模型 gcForest 的不足，并基于这些不足进行了改进。改进主要分为三方面：(1) 使用带类别权重的随机森林，增加集成方法的多样性，并缓解特征淹没问题；(2) 使用序列化训练的随机森林，使得提取到的特征更加互补；(3) 使用层级间特征选择机制，剔除冗余特征，保留有用特征，便于提取更高阶的特征，加速模型收敛。在本章末尾给出了改进方法的整体框架和算法流程。下一章中我们将通过实验来验证该方法的实际效果。

第4章 实验设计与结果分析

为了验证本文提出的改进方法在基于表面肌电信号的手势分类问题中可以取得较好的表现,且该方法在其它分类问题中也可以取得不错的效果,本章共进行了三组实验,前两组实验使用表面肌电信号的手势分类数据集,第三组实验使用了其它的若干公开数据集。各实验均包括数据集简介、对比实验及实验参数设置、实验结果和结论分析。

4.1 数据集简介

本文中采用的数据集是 UCI 机器学习库中的基于 sEMG 的手势分类数据集^[4]。数据采集过程要求 5 名年龄相仿的健康受试者重复以下 6 个基本动作:

- (a) 圆柱形抓握: 用于抓圆柱形工具
- (b) 钩形抓握: 用来支撑重物
- (c) 指侧抓握: 用来夹持薄而平的物体
- (d) 指腹抓握: 手掌面向物体的抓取
- (e) 球形抓握: 用于夹持球形工具
- (f) 指尖抓握: 用于拿小工具

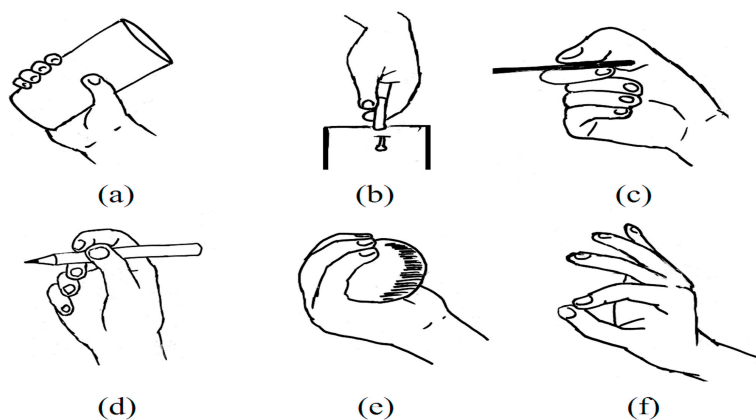


图 4.1 六个手势示意图^[8]

对每个受试者，要求每个基本动作执行 6 秒，每个基本动作重复做 30 次。每位受试者共记录 180 个 6 秒长的 2 通道表面肌电信号（2 通道即使用两个电极同时采集信号，最终得到两个表面肌电信号）。2 个通道记录的数据共同记录了该动作的相关信息。

具体来说，信号数据以 500HZ 的采样率进行采集（即每秒采集 500 个信号点），并进行带通滤波，使用 Butterworth 带通滤波器（上下阈值分别为 500Hz 和 15Hz）和陷波滤波器（50Hz）对采集到的表面肌电信号数据进行处理，以消除线路干扰和噪声。

采集到的表面肌电信号示例如图 4.1 所示：

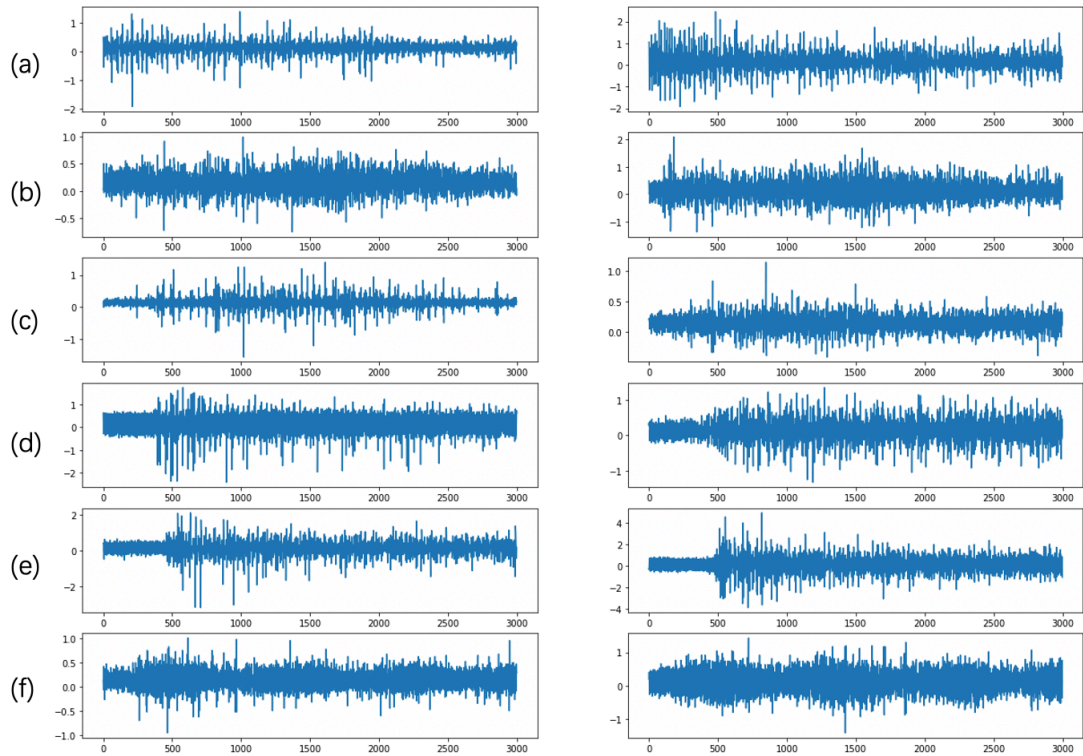


图 4.2 六个手势的 2 通道表面肌电信号示意图

可以看到，对每个受试者的每个动作，我们都可以得到 2 个 3000 维的特征向量，若将其拼接在一起视为 1 个样本，则特征维数为 6000 维，而我们一共有 900 个样本，6 个动作类别，每类动作只有 150 个样本。因此我们面临的是一个典型的小样本、高维度数据上的多分类问题。

4.2 基于 sEMG 进行手势分类问题的实验设计

本节将进行基于 sEMG 的手势分类的实验，以此来验证本文提出的改进方法的有效性。

首先我们将直接使用原始的 sEMG 数据作为输入，将本文提出的改进方法与 gcForest 方法以及经典的机器学习方法的结果进行比较，以此说明改进的方法对于原始 sEMG 信号这种高维数据具有较强的特征提取和转换能力，能得到较好的表现。

接下来我们使用人工提取的特征作为输入。具体来说，我们从原始 sEMG 信号中提取出频谱图特征作为频域特征，并使用 1D-LBP 方法提取出 LBP 直方图特征作为时域特征，将两者拼接起来得到模型输入，然后使用本文方法中的级联森林（不再使用多粒度扫描）部分进行实验，与 gcForest 及其它经典的机器学习方法的结果进行比较，并与以往的研究结果进行比较，以此说明改进的方法在基于 sEMG 进行手势分类的问题中可以取得有竞争力的表现。

4.2.1 使用原始信号作为输入

如 4.1 节所述，原始数据中每个样本（即一个受试者所做的一次动作）包含 2 个 3000 维的特征向量。这里我们简单地把两个向量拼接到一起得到一个 6000 维的特征向量作为模型的输入。此时问题变成了一个样本数 900、特征维数 6000 的 6 分类问题，要想解决该问题，传统的做法是人工提取一些特征，然后再用经典的机器学习算法解决问题。深度学习受制于样本数量，通常容易产生过拟合。而深度森林方法对于此类问题的处理有其独特优势，一方面深度森林可以自动确定层数以使得模型复杂度适应数据集，另一方面深度森林具有较强的特征转换能力，可以提取高阶特征用于模型的训练。

为了说明本文提出的改进的深度森林方法在处理此类小样本高维数据的多分类问题时有较大优势，接下来的实验中我们将把改进的方法与经典的机器学习方法（包括 K 近邻、支持向量机、随机森林、多层感知机、卷积神经网络）以及原深度森林方法 gcForest 进行比较。本实验中我们采用分类精度作为模型表现的衡量指标。每次实验都随机取 2/3 的数据作为训练集，剩下 1/3 的数据作为测试集，重复进行 20 次实验并将分类精度的平均值作为模型的最终表现。

实验中使用的各分类器的参数设置（所有默认设置均为 sklearn 库中参数的默认设置）如下：（1）K 近邻分类器（KNN）：K=7，其它参数使用默认设置；（2）支持向量机（SVM）：采用高斯核函数，其它参数使用默认设置；（3）随机森林（RF）：基决策树数量设为 600，结点划分时随机选取的特征子集大小为 \sqrt{m} (m 为总特征数)，其它参数使用默认设置；（4）多层感知机（MLP）：采用 $input - 1024 - 512 - 256 - output$ 的网络机构；（5）卷积神经网络（CNN）：使用 32 个大小为 5 的一维卷积核构造卷积层，卷积层后连接一个池化层，重复上述结构两次，然后是 128 个神经元的全连接层，激活函数为 relu，最后是 6 个神经元的输出层，激活函数为 softmax；（6）深度森林（gcForest）：每层 2 个含 600 棵树的随机森林和 2 个含 600 棵树的完全随机森林，多粒度扫描的窗口大小为 20，步长为 20；（7）改进的深度森林：每层 6 个含 200 棵树的随机森林和 2 个含 600 棵树的随机森林，多粒度扫描的窗口大小分别为 20，步长为 20， $\alpha = 32$ ， $\beta = \sqrt{m}$ 。表 4.1 列出了各分类器的参数设置。需要注意，改进方法中使用含 200 棵树的随机森林是为了减少计算开销，深度森林使用含 600 棵树的随机森林则是为了保证模型效果，若两者使用完全相同的参数设置，则改进方法优势更加明显。

表 4.1 各分类器的超参数设置

KNN	K = 5
SVM	kernel=rbf
RF	基决策树的数量为 600
MLP	网络结构为 input-1024-512-256-output
CNN	网络结构为 input-32 × 5conv1d-pooling-32 × 5conv1d-pooling-32 × 5conv1d-pooling-128(relu)-6(softmax)-output
gcForest	每层 2 个含 600 棵树的随机森林和 2 个含 600 棵树的完全随机森林 多粒度扫描的窗口大小分别为 20，步长为 20
Proposed method	每层 6 个含 200 棵树的随机森林和 2 个含 600 棵树的随机森林 多粒度扫描的窗口大小分别为 20，步长为 20， $\alpha = 32, \beta = \sqrt{m}$

实验结果如表 4.2 所示。K 近邻、支持向量机、随机森林等经典的机器学习算法的分类精度分别为 19.50%、35.33%、44.16%，其中随机森林的表现最好。神经

网络模型中，多层感知机由于难以有效提取高维序列的局部特征，因此表现较差，精度只有 28.02%。卷积神经网络则取得了较好的表现，分类精度达到了 58.93%。相比之下，gcForest 模型分类精度达到了 66.77%，超过了上述经典机器学习方法以及深度学习方法。而我们提出的改进的深度森林方法则取得了 73.53% 的分类精度，超过了上述经典机器学习方法以及深度学习方法，且比原 gcForest 模型的表现提升了 20.34%。

表 4.2 各分类器实验结果比较

KNN	19.50%
SVM	35.33%
RF	44.16%
MLP	28.02%
CNN	58.93%
gcForest	66.77%
Proposed method	73.53%

使用如此高维的原始信号特征作为模型输入进行实验的目的并不是得到最佳的手势分类结果，而是为了验证深度森林模型具有通过特征转换提取高阶特征的能力以及处理此类小样本高维数据的优势。该实验能达成以上目的是因为原始信号较为杂乱且包含噪声，而且动作的起始时间和结束时间并不固定，这意味着模型要自动从高维的杂乱信号中提取并整合有用的局部特征信息，才能取得较好的分类效果。从实验结果来看，无论是 gcForest 模型，还是改进后的深度森林模型，在以高维的原始信号作为输入的情况下都取得了优于传统机器学习模型和神经网络模型的表现。而我们提出的改进后的深度森林模型相比原 gcForest 模型的表现又有了较大幅度的提升，这说明我们提出的几点改进措施是有效的，改进后的方法确实可以更加充分地从原始特征中提取特征信息，从而在小样本高维数据上的多分类问题中取得更好的表现。

下一节中，我们将对原始特征做简单的特征提取处理，然后将提取到的特征输入到各分类器中进行实验，并通过比较实验结果来进一步验证本文提出的改进方法的有效性。

4.2.2 使用提取的特征作为输入

上一节中我们直接把原始信号作为模型的输入，得到的模型表现并不理想，这主要是因为原始的表面肌电信号维度很高、包含噪声且动作的起始时间不统一，因此对模型提取局部特征的能力是一个很大的挑战。深度森林模型对局部特征的提取和整合主要依靠多粒度扫描来完成，而本节中我们将对原始表面肌电信号进行预处理，通过简单的方法人工提取时域特征和频域特征来代替多粒度扫描的工作。

由于采集数据时受试者自己来控制完成动作的速度和力量，因此采集到的数据存在个体差异，为了提取出更能反映动作本质的特征，同时又不引入过多的人工来构造特征，我们先对原始信号做简单的特征提取。具体分为两个部分：（1）计算原始信号数据的功率谱密度，以此代表信号的频域信息；（2）计算原始信号数据的局部二值模式（LBP）直方图，以此代表信号的时域信息。需要注意的是，这里我们并没有直接提取常用的高阶表面肌电信号特征^[7]，而是简单地将信号在频域和时域以另一种形式表示出来，可以认为我们提取出的特征是低阶特征。如表 4.3 所示。

表 4.3 表面肌电信号的时频域特征

	常用特征	本文采用特征
时域特征	Zero crossing Slope sign change Wilson amplitude Waveform length Integrated EMG Mean absolute value Variance Root-mean-square	LBP histogram
频域特征	Mean frequency Median frequency Peak frequency	Power spectral density

下面具体说一下我们采用的特征提取方法。功率谱密度 (Power Spectral Density, PSD)，实际就是通过一定方法求解信号的功率随频率变化的曲线。我们采用 Welch 方法^[34] 来计算功率谱密度，Welch 方法的思路是：先把信号分成多段（各段之间可以有重叠），然后把窗口函数（Hamming、Hanning 等）加到每一段信号上，求出每一段信号的功率谱，最后对每段数据的功率谱进行平均，得到整个信号的功率谱。相比传统的快速傅里叶变换（FFT）方法，Welch 方法的分辨率更高，在脑电信号、眼电信号、肌电信号等生物信号的处理中有着广泛的应用。功率谱密度表示了信号功率随着频率的变化情况，或者说信号功率在频域的分布状况，因此可以将其视作信号的频域特征。下面给出了一个原始信号及其对应的功率谱密度的图示：

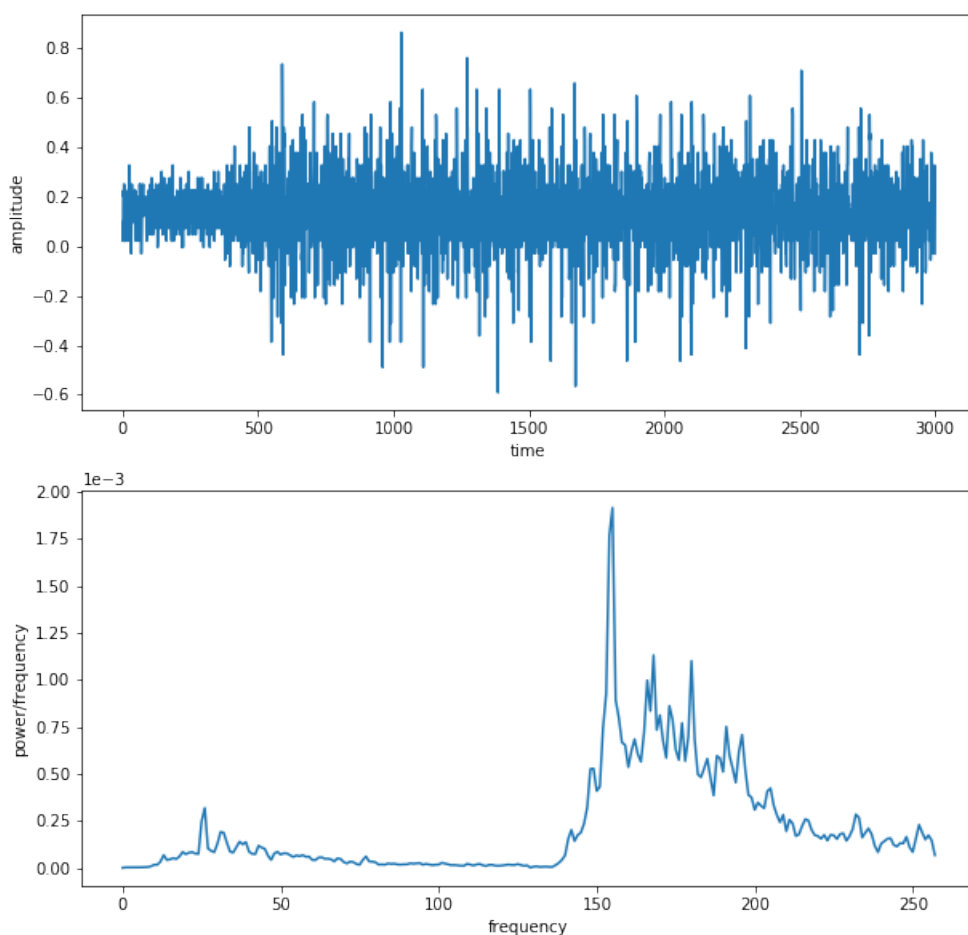


图 4.3 原始信号和对应的功率频谱图

使用 Welch 方法对每个样本包含的两个 3000 维的原始信号分别提取出一个 129 维的 PSD 特征，然后拼接在一起得到一个 258 维的特征，作为每个样本对应的频域特征。

除了以上频域特征的提取，本文还提取了时域特征。和常用的信号平均振幅、信号下面积、信号过零率等时域特征不同，我们这里主要使用信号的一维局部二值模式 (1D-LBP) 特征，实际上之前已经有基于 LBP 的表面肌电手势识别研究^[7]。我们这里使用 LBP 直方图特征的原因在于，我们的数据来自多个不同的受试者，受试者之间无论是肌肉力量还是动作速度都存在个体差异，因此直接使用与振幅大小相关的特征或与振幅变化速度相关的特征都不是十分合理，而 LBP 作为一种提取纹理模式的方法则可以有效地避免个体差异对手势识别的干扰，下图给出了一个简单的示例来说明这一点。

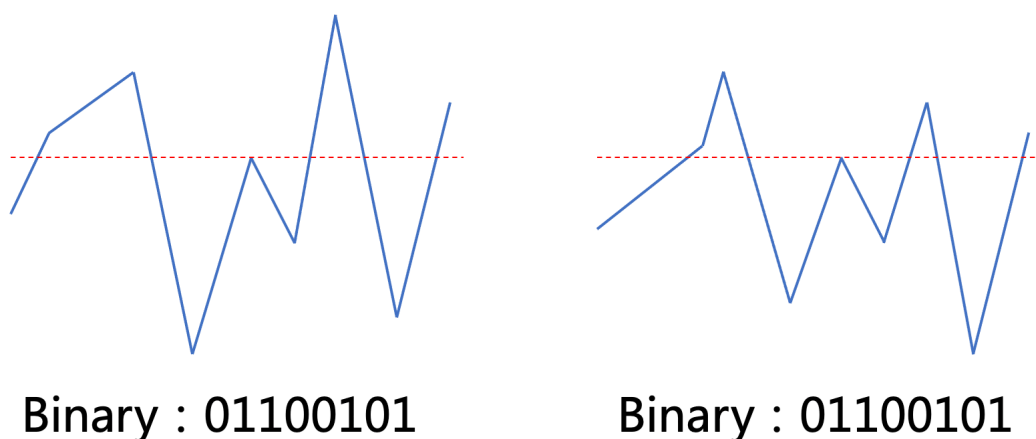
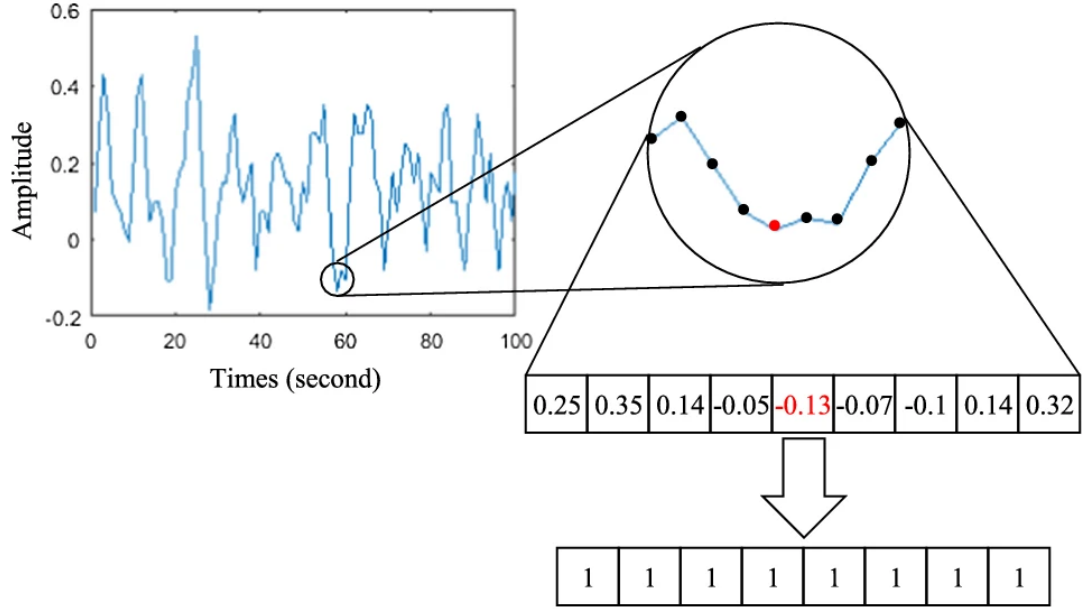


图 4.4 两个不同的信号对应相同的 LBP 值

在 2.4 小节中我们已经简单地介绍了局部二值模式 (LBP) 以及一维局部二值模式 (1D-LBP) 算法的原理和计算过程，在提取 sEMG 的 1D-LBP 特征时，我们使用一个长度为 9 的窗口沿整个信号滑动，在每个窗口内都计算一个 LBP 值，如图 4.5 所示。最后将整个信号的所有 LBP 值计数统计得到一个直方图，该直方图就可以作为整个信号波形的纹理表征，可以作为原始信号的时域特征使用。需要注意的是，如果原始信号的长度不一致，在使用 LBP 直方图的时候需要先进行归一化，在本文所使用的数据集中，每个样本均包含两个 3000 维的信号数

据, 因此可以直接对这两个 3000 维的信号分别提取一个 256 维的 LBP 直方图特征, 然后拼接在一起得到一个 512 维的特征向量, 不需要做归一化。



$$1DLBP_8^{All} = 1 \times 1 + 1 \times 2 + 1 \times 4 + 1 \times 8 + 1 \times 16 + 1 \times 32 + 1 \times 64 + 1 \times 128 = 255$$

图 4.5 从 sEMG 原始信号中提取 LBP^[7]

按上述步骤可以从原始的 2 通道表面肌电信号中分别提取出 258 维的频域特征和 512 维的时域特征, 将两者拼接在一起就得到了一个 770 维的特征向量作为模型的输入, 相比原始数据的 6000 维特征维度降低了约一个数量级。当然最关键的是各样本提取后的特征具有相同的物理意义, 因此在训练过程中无需再从特征的不同位置提取局部特征进行整合, 大大降低了训练难度, 同时也节约了计算资源。

接下来要进行的实验中我们依然使用与 4.2.1 节中相同的分类器, 除了 gcForest 以及本文提出的改进方法外, 其它分类器的超参数设置与表 4.1 相同。gcForest 采用每层 2 个含 600 棵树的随机森林和 2 个含 600 棵树的完全随机森林; 本文提出改进的深度森林方法采用每层 6 个含 200 棵树的随机森林和 2 个含 600 棵树的随机森林, $\alpha = 32$, $\beta = \sqrt{m}$ 。gcForest 方法以及本文提出的改进方法均不再使用多粒度扫描。

实验结果如表 4.4 所示。可以看到，在使用提取后的特征进行手势分类时，各分类器的表现都比使用原始信号作为输入有了大幅提高。在经典的机器学习方法中，K 近邻分类器和随机森林分类器的分类精度都达到了 95% 以上，其中，K 近邻的分类精度为 95.7%，随机森林的分类精度为 96.85%。支持向量机的分类精度较低，为 80.00%。在神经网络模型中，多层感知机和卷积神经网络的表现都不是很理想，分类精度分别为 85.56% 和 69.83%。相比之下，gcForest 模型分类精度达到了 97.16%，超过了上述经典机器学习方法以及深度学习方法。而改进的深度森林方法则取得了 98.06% 的分类精度，比原 gcForest 模型的表现提升了 31.69%。

表 4.4 各分类器实验结果比较——手势分类

KNN	95.7%
SVM	80.00%
RF	96.85%
MLP	85.56%
CNN	69.83%
gcForest	97.16%
Proposed method	98.06%

上述实验结果表明了改进的方法在基于表面肌电信号的手势分类问题中具有比经典机器学习方法以及深度学习方法更好的表现，且相比 gcForest 模型的表现也有较大提升。

下面我们将针对各个受试者的信号数据分别进行训练，并记录各分类器在各个受试者数据集上的分类精度。进行这样一个实验的原因有三点：（1）在实际应用中，若我们能通过单独受试者的少量表面肌电信号数据训练出精度较高的模型，则有利于在不引入额外样本的情况下制造个性化的肌电假肢；（2）样本更少的分类任务更具挑战性，可以进一步比较各分类器在小样本分类问题中的表现；（3）之前很多有关 sEMG 手势分类的研究中的实验是以受试者为单位进行的，为了方便将改进的方法与之前的研究结果进行比较，因此我们也进行一组针对各个受试者的实验。

如 4.1 节所述, 实验采用的表面肌电信号数据集来自 5 名年龄相仿 (20~22 岁) 的健康受试者 (2 男 3 女), 现按照他们在数据集中的排列顺序将其依次命名为受试者 1~5, 每个受试者对应数据集的样本为 180, 6 个手势各有 30 个样本, 特征维数为 770。采用 gcForest 模型和改进的方法在各受试者的数据集上进行实验, 并将实验结果与之前已有研究的实验结果进行比较, 结论如表 4.5 所示。

表 4.5 与之前研究结果的比较——手势分类

Researchers (year)	Feature extraction	Method	Classification accuracy (%)					
			#1	#2	#3	#4	#5	AVG
Sapsanis(13) ^[4]	EMD	linear classifier	87.25	88.05	85.53	90.42	94.80	89.21
Ruangpaisarn(15) ^[30]	SVD	SMO classifier	96.67	98.89	96.67	98.89	100.00	98.22
Iqbal(17) ^[5]	SVD & PCA	KNN	82.78	87.67	83.11	90.00	90.00	86.71
Akben(17) ^[6]	Histogram & energy	K-means	93.04	86.66	97.00	99.23	97.66	94.72
Subasi(18) ^[9]	WPD	rotation forest	95.56	88.88	92.22	92.22	98.33	93.44
Nishad(19) ^[8]	TQWT-FB	KNN	98.33	97.78	99.44	98.89	98.33	98.55
Yavuz(19) ^[32]	Cepstral coefficients	GRNN	98.33	99.48	99.44	99.78	99.69	99.34
Z-H.Zhou(17) ^[18]	PSD & LBP	<i>gcForest</i>	99.72	98.88	99.00	98.66	99.33	99.12
Proposed method	PSD & LBP	<i>gcForest</i> *	100.00	99.44	99.16	98.83	99.49	99.38

表中选取了几项关于 sEMG 手势分类的有代表性的研究, 并按年份排序。这些研究都使用相同的数据集, 即 UCI 机器学习库中的基于表面肌电信号的手势分类数据集^[4]。Sapsanis 等人在 2013 年创建了我们实验中使用的原始数据集^[4], 提出了一种基于 EMD 的特征提取方法, 并使用线性分类器对手势进行了分类。该方法在五名受试者的数据集上平均精度为 89.21%。Ruangpaisarn 和 Jaiyen^[30]在 2015 年提出了一种基于 SVD-SMO 分类器的方法, 平均准确率达到 98.22%。Iqbal 等人^[5]于 2017 年提出利用 SVD 和 PCA 方法进行特征提取, 并使用 KNN 进行分类的方法取得了 86.71% 的平均精度。同年, Akben^[6]提出一种基于直方图和能量的 K-means 聚类方法, 平均精度为 94.72%。2018 年, Subasi 等人^[9]引入了一种基于小波包分解 (WPD) 的特征提取方法, 并使用旋转森林作为分类器, 得到了 93.44% 的平均精度。2019 年, Nishad 等人^[8]使用 TQWT-FB 的特征提取方法, 并使用 KNN 作为分类器, 平均精度为 98.55%。同年, Yavuz 等人^[32]将倒频谱分析与 GRNN 分类器相结合, 平均精度为 99.34%。这是在已有文献中找到的最佳分类结果。而本文中我们基于功率谱密度 (PSD) 和局部二值模式

(LBP) 特征, 使用 *gcForest* 模型以及改进的 *gcForest** 模型分别取得了 99.12% 和 99.38% 的平均分类精度。与以往的研究相比, 深度森林方法得到了较好的分类表现。其中改进的 *gcForest** 模型的表现优于表中提到的所有研究。这说明我们提出的改进的深度森林方法在基于表面肌电信号的手势分类问题中可以取得有竞争力的表现。

需要注意的是, 虽然改进方法的表现比现有最好结果的提升只有约 6%, 但我们的特征工程相比其它研究要简单很多, 也没有引入复杂的专家知识, 因此能取得 99.38% 的平均分类精度已经是不错的结果, 这说明了改进方法具有从高维小样本数据中自动提取有效高阶特征并依此进行准确分类的能力。

4.3 基于其它数据集的实验设计

在 4.2 节中, 我们将改进的 *gcForest* 方法应用于基于表面肌电信号的手势分类问题中, 并取得了较好的结果。为了进一步说明改进的方法具有一定的普适性, 即可以在多场景的分类任务中取得较好的效果, 本节将基于其它的一些公开数据集进行实验。

音乐分类——GTZAN

GTZAN 数据集^[33] 包含 10 种类型的音乐片段, 每一种音乐有 100 段 30 秒长的音频数据。我们使用梅尔倒谱系数 (Mel-scale Frequency Cepstral Coefficients, 简称 MFCC) 作为各段音频的特征, 对原始音频数据进行处理后每个样本特征维度为 20×1400 。我们将数据集划分为训练集和测试集, 每次随机抽取 700 个样本用于训练, 剩下 300 个样本进行测试。重复进行 20 次实验取结果的平均值作为模型的最终结果。

在本实验中, 我们依然使用 4.2 节中的几种分类器与 *gcForest* 模型及本文提出的模型进行比较。实验中使用的各分类器的参数设置 (所有默认设置均为 *sklearn* 库中参数的默认设置) 如下: (1) K 近邻分类器 (KNN): $K=7$, 其它参数使用默认设置; (2) 支持向量机 (SVM): 采用高斯核函数, 其它参数使用默认设置; (3) 随机森林 (RF): 基决策树数量设为 600, 结点划分时随机选取

的特征子集大小为 \sqrt{m} (m 为总特征数), 其它参数使用默认设置; (4) 多层感知机 (MLP): 采用 $input - 1024 - 256 - output$ 的网络机构; (5) 卷积神经网络 (CNN): 使用 32 个大小为 10 的卷积核构造卷积层, 后接池化层, 然后是结构为 1024-512 的全连接层, 激活函数为 relu, 最后是 10 个神经元的输出层, 激活函数为 softmax; (6) 深度森林 (gcForest): 每层 2 个含 600 棵树的随机森林和 2 个含 600 棵树的完全随机森林, 多粒度扫描的窗口大小为 20, 步长为 1; (7) 改进的深度森林: 每层 6 个含 200 棵树的随机森林和 2 个含 600 棵树的随机森林, 多粒度扫描的窗口大小为 20, 步长为 1, $\alpha = 32$, $\beta = \sqrt{m}$ 。

实验结果如表 4.6 所示。可以看到, 在使用提取的 MFCC 特征进行音乐类型分类时, 经典的机器学习方法表现都不是很好, K 近邻分类器的平均分类精度最低, 仅为 29.67%。支持向量机和随机森林的分类精度则分别为 47.00% 和 41.86%。相比之下, 神经网络模型的表现整体优于上述方法, 多层感知机模型和卷积神经网络模型分别取得了 51.83% 和 58.26% 的平均分类精度。深度森林方法在音乐分类的任务中表现最好, 其中, gcForest 模型分类精度达到了 61.99%, 优于上述经典机器学习方法以及深度学习方法。而改进的深度森林方法则取得了 64.67% 的分类精度, 比原 gcForest 模型的表现提升了 7.05%。

表 4.6 各分类器实验结果比较——音乐分类

KNN	29.67%
SVM	47.00%
RF	41.86%
MLP	51.83%
CNN	58.26%
gcForest	61.99%
Proposed method	64.67%

情绪识别——IMDB

IMDB 数据集^[35]来自 IMDB 网站的 25000 条电影评论, 按情绪 (正面/负

面)进行标记。我们提取各评论的 TF-IDF 特征,并将其用于模型训练。由于原数据集规模较大,且提取出的特征维数很高(数量级为 10^4),因此出于时间和实验设备的限制,我们这里随机从正负样本中各抽取 2000 个,构成一个总大小为 4000 的数据集,以此作为我们的实验数据集,基于此数据集提取的特征维数为 24108。在该实验中,因为我们使用的特征不具有空间和位置特性,因此不再使用 CNN 进行实验,在 gcForest 以及改进的深度森林方法中也不再使用多粒度扫描。其余分类器参数设置同上。

实验结果如表 4.7 所示。可以看到,在使用提取的高维 TF-IDF 特征进行情绪识别时,在经典的机器学习方法中,K 近邻分类器和支持向量机的表现都不是很好,两者的分类精度分别为 71.67% 和 68.58%。随机森林表现则较好,取得了 83.67% 的分类精度。多层感知机模型表现优于以上经典分类器,平均分类精度为 84.72%。相比之下,gcForest 模型分类精度与多层感知机模型基本持平,为 84.94%。而改进的深度森林方法则取得了 85.33% 的分类精度,比原 gcForest 模型的表现提升了 2.59%。

表 4.7 各分类器实验结果比较——情绪识别

KNN	71.67%
SVM	68.58%
RF	83.67%
MLP	84.72%
gcForest	84.94%
Proposed method	85.33%

新闻分类——NEWS20

NEWS20 数据集包含关于 20 个主题的大约 18000 篇新闻文章,这些文章被分成两个子集:一个用于训练,另一个用于测试。训练集和测试集之间的划分基于一个特定日期,该日期前发表的文章为训练集,该日期后为测试集。我们提取各评论的 TF-IDF 特征,并将其用于模型训练。由于原数据集规模较大,且

提取出的特征维数很高（数量级为 10^4 ），因此出于时间和实验设备的限制，我们这里从 20 个新闻主题中选择 4 个，分别为：alt.atheism、talk.religion.misc、comp.graphics、sci.space，构成一个总大小为 2034 的数据集，以此作为我们的实验数据集，基于此数据集提取的特征维数为 34118。在该实验中，因为我们使用的特征不具有空间和位置特性，因此不再使用 CNN 进行实验，在 gcForest 以及改进的方法中也不再使用多粒度扫描。其余分类器参数设置同上。

实验结果如表 4.8 所示。可以看到，在使用提取的 TF-IDF 特征进行新闻分类时，支持向量机的平均分类精度较低，为 87.13%。K 近邻和随机森林的表现稍好，分类精度分别为 88.32% 和 88.38%。多层感知机模型表现优于以上经典分类器，平均分类精度为 91.38%。相比之下，gcForest 模型分类精度为 90.99%，优于上述经典机器学习方法，但相比多层感知机模型的结果略差。而改进的深度森林方法则取得了 94.05% 的分类精度，在所有分类器中表现最好，比原 gcForest 模型的表现提升了 33.96%。

表 4.8 各分类器实验结果比较——新闻分类

KNN	88.32%
SVM	87.13%
RF	88.38%
MLP	91.38%
gcForest	90.99%
Proposed method	94.05%

DNA 分类——Splice-junction Gene Sequences Data Set

Splice-junction Gene Sequences Data Set^[36] 关注的问题是，给定一个 DNA 序列，识别外显子（剪接后 DNA 序列保留的部分）和内含子（剪接后 DNA 序列的部分）之间的边界。数据集中的 DNA 片段共可以分为三个类别，DNA 片段以“CCAGCTGCAT……”的数据格式保存。在解决该问题时，我们先对原始特征进行处理，得到 287 维的特征，作为实验的输入。值得注意的是，原始数据集中共

包括 3190 个样本，而由于我们特征维数并不是非常高，因此采用传统的机器学习方法（如支持向量机、随机森林等）已经可以取得较好的效果，为了考查各分类器从小样本中提取特征进行分类的能力，我们在实验中每次都随机抽取 20% 的数据（即 638 个样本）作为实验数据集，然后将其中 50% 用作训练集，剩下 50% 用作测试集。重复进行 10 次实验取平均精度作为最终的模型表现。

实验结果如表 4.9 所示。可以看到，K 近邻和支持向量机的平均分类精度较低，分别为 75.58% 和 86.87%。随机森林和多层感知机模型表现较好，平均分类精度分别为 93.70% 和 91.63%。相比之下，gcForest 模型分类精度为 94.95%，优于上述经典机器学习方法。而改进的深度森林方法则取得了 95.49% 的分类精度，在所有分类器中表现最好，比原 gcForest 模型的表现提升了 10.69%。

表 4.9 各分类器实验结果比较——DNA 分类

KNN	75.58%
SVM	86.87%
RF	93.70%
MLP	91.63%
gcForest	94.95%
Proposed method	95.49%

在本节中，我们基于其它的一些公开数据集进行实验，比较了本文提出的改进方法和几种经典的机器学习算法以及神经网络模型在不同场景分类任务中的表现。实验结果表明，本文提出的方法不仅适用于基于表面肌电信号的手势分类问题，在多场景的分类任务问题中都能取得良好的表现，在小样本高维数据上的多分类问题中的优势则更加明显。

4.4 本章小结

在本章中，我们主要进行了三组实验。第一组实验将原始 sEMG 信号作为输入，比较各模型的效果，改进的方法优于经典分类器及 gcForest 模型，在所有分类器中取得了最好的表现，说明了改进的方法具有较强的从高维数据中提取特征

的能力。第二组实验使用 PSD 和 LBP 对 sEMG 原始信号进行特征提取，将提取后的特征作为输入，比较各模型效果，改进的方法优于经典分类器及 gcForest 模型，在所有分类器中取得了最好的表现。之后我们又将实验结果与之前的研究结果进行比较，说明了改进的方法在 sEMG 手势分类问题中取得了有竞争力的表现。第三组实验比较了各模型在多个领域公开数据集上的表现，证明了改进的方法在多场景的分类任务中均可以取得良好的表现，且在高维小样本数据上的多分类问题中优势尤为明显。

第5章 总结与展望

表面肌电信号在人机交互中发挥着重要作用，其中基于表面肌电信号的肌电假肢是很有前景的应用，因此基于表面肌电信号的手势识别问题作为肌电假肢领域的重要问题得到了研究者的广泛关注。

本文针对基于表面肌电信号的手势分类问题提出了新的解决方法，该方法在深度森林模型 gcForest 的基础上进行了三点改进：(1) 在级联森林的每一层中使用带类别权重的随机森林；(2) 在级联森林的每一层中使用序列化训练的随机森林；(3) 使用层级间特征选择机制。经过实验验证，改进的方法在基于表面肌电信号的手势分类问题中取得了具有竞争力的表现。此外，为了验证改进的方法具有一定的普适性，我们又在其它几个公开数据集上验证了模型效果。

虽然改进的深度森林方法在基于表面肌电信号的手势识别这一类小样本高维度数据上的多分类问题中具有良好的表现，但同时我们的算法也有不够完善的地方，在未来的工作中我们需要解决的问题如下：

(1) 算法的复杂度相比原 gcForest 较高。改进后的深度森林方法相比原来的深度森林方法复杂度有所提高，这主要是因为序列化训练的随机森林以及额外设置的用于检验模型精度和进行特征选择的随机森林导致的。在本文的实验中，深度森林模型的用时通常比随机森林高一个数量级（深度森林数十秒，随机森林数秒）。针对这一问题，后续我们可以考虑加入目前已有研究中的一些机制，比如 M.Pang 等人^[37]提出的置信度筛选机制，即无需让所有样本通过所有层，若样本在某一层的分类置信度超过某个阈值，则可以直接输出。又比如 Shiwen Ni 等人^[38]提出的在多粒度扫描中加入 pooling 机制等。

(2) 对于类别数很多的分类问题计算开销太大。因为我们对每个类别都训练一个随机森林，所以当类别数较多的时候带来的计算开销会很大。可能的改进方案是先利用层次聚类等方法将原始类别划分为几个子集，每个子集中包含的类别较为相似，然后对各类别子集分别训练一个随机森林。

同时，由于时间和条件的限制，本文提出的对深度森林算法的改进并不全面，未来可以尝试的改进方法包括：

(1) 当前的深度森林算法每一层的训练均以最终分类标签为目标, 这可能会导致模型过早收敛, 无法充分利用层级结构的优势来提取特征。受到 Hinton^[39] 提出的知识蒸馏方法的启发, 考虑是否可以启发式的计算出样本的 soft label, 并将其用于深度森林前几层的训练。

(2) 对深度森林中使用的随机森林进行改进。目前我们使用的深度森林对样本的类别概率预测是各棵树结果的简单平均, 而实际上各棵树的表现有好有坏, 我们应当给予表现较好的树较大的权重。目前已经有研究提出了相关的方法, 比如 V.Utkin 等人^[40] 就提出了一种对随机森林中的树进行加权的方法。

(3) 对深度森林的内存开销进行优化。深度森林目前无法用于较大规模数据集对原因就在于它的内存开销很大, 因此为了将其应用到更多对问题和场景中去, 内存的优化很有必要。这不仅需要利用数据结构与算法的知识对算法本身进行高效实现, 也需要对算法本身进行优化。

参考文献

- [1] P. Xiong, J. Tang. Surface emg signal preamplifier system [J]. *Computer Engineering and Applications*, 51(1):218–222, 2015.
- [2] N. Wang, K. Lao, X. Zhang, J. Lin, and X. Zhang. The recognition of grasping force using LDA. *Biomedical Signal Processing and Control*, 47:393–400, 2019.
- [3] 罗永昭, 孙为. 建国以来我国假肢的发展 [J]. *中国矫形外科杂志*, 17(17):1325–1328, 2009
- [4] C. Sapsanis, G. Georgoulas, A. Tzes, D. Lymberopoulos. Improving emg based classification of basic hand movements using emd. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5754–5757. IEEE, 2013.
- [5] O. Iqbal, S. Fattah, S. Zahin. Hand movement recognition based on singular value decomposition of surface emg signal. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pages 837–842. IEEE, 2017.
- [6] S. Akben. Low-cost and easy-to-use grasp classification, using a simple 2-channel surface electromyography (semg). *Biomedical Research*, 28:577–582, 04 2017.
- [7] S. Tabatabaei, A. Chalechale. Local binary patterns for noise-tolerant semg classification. *Signal, Image and Video Processing*, 13(3):491–498, 2019.
- [8] A. Nishad, A. Upadhyay, RB. Pachori, UR. Acharya. Automated classification of hand movements using tunable-q wavelet transform based filter-bank with surface electromyogram signals. *Future Generation Computer Systems*, 93:96–110, 2019.
- [9] A. Subasi, L. Alharbi, R. Madani, S. Qaisar. Surface emg based classification of basic hand movements using rotation forest. In *2018 Advances in Science and Engineering Technology International Conferences (ASET)*, pages 1–5. IEEE, 2018.
- [10] N. Sikder, A. Arif, A. Nahid. Heterogeneous hand guise classification based on surface electromyographic signals using multichannel convolutional neural network. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2019.
- [11] F. Ayaz, A. Ari, D. Hanbay. Classification of emg signals by lrf-elm. In *2017 Interna-*

- tional Artificial Intelligence and Data Processing Symposium (IDAP)*, pages 1–6. IEEE, 2017.
- [12] C. Song, C. Chen, Y. Li, X. Wu. Deep reinforcement learning apply in electromyography data classification. In *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pages 505–510. IEEE, 2018.
- [13] N. Rabin, M. Kahlon, S. Malayev, A. Ratnovsky. Classification of human hand movements based on emg signals using nonlinear dimensionality reduction and data fusion techniques. *Expert Systems with Applications*, 149:113281, 2020.
- [14] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [15] 周志华. 机器学习 [J]. 北京: 清华大学出版社, 2016.
- [16] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [17] Y. Bengio, N. Boulanger-Lewandowski, R. Pascanu. Advances in optimizing recurrent networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8624–8628. IEEE, 2013.
- [18] Z. Zhou, J. Feng. Deep forest. *National Science Review*, 6(1):74–86, 2019.
- [19] A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [20] A. Graves, A. Mohamed, G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.
- [21] L. Breiman. Randomizing outputs to increase prediction accuracy. *Machine Learning*, 40(3):229–242, 2000.
- [22] TK. Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.
- [23] Y. Guo, S. Liu, Z. Li, X. Shang. BCDForest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. *BMC bioinformatics*, 19(5):118, 2018.
- [24] X. Zeng, Y. Zhong, W. Lin, Q. Zou. Predicting disease-associated circular rnas using deep forests combined with positive-unlabeled learning methods. *Briefings in bioinformatics*, 21(4):1425–1436, 2020.
- [25] T. Ojala, M. Pietikainen, T. Maenpaa. Multiresolution gray-scale and rotation invariant

- texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.
- [26] Yı. Kaya, M. Uyar, R. Tekin, S. Yıldırım. 1D-local binary pattern based feature extraction for classification of epileptic eeg signals. *Applied Mathematics and Computation*, 243:209–219, 2014.
- [27] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, J. Candela. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pages 1–9, 2014.
- [28] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [29] M. Raez, M. Hussain, F. Mohd-Yasin. Techniques of emg signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8(1):11–35, 2006.
- [30] Y. Ruangpaisarn, S. Jaiyen. SEMG signal classification using smo algorithm and singular value decomposition. In *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 46–50. IEEE, 2015.
- [31] N. Kakoty, S. Hazarika. Recognition of grasp types through principal components of dwf based emg features. In *2011 IEEE International Conference on Rehabilitation Robotics*, pages 1–6. IEEE, 2011.
- [32] E. Yavuz, C. Eyupoglu. A cepstrum analysis-based classification method for hand movement surface emg signals. *Medical & biological engineering & computing*, 57(10):2179–2201, 2019.
- [33] G. Tzanetakis, P. Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.
- [34] P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.
- [35] A. Maas, R. Daly, P. Pham, D. Huang, A. Ng, C. Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142–150, 2011.

- [36] M. Noordewier, G. Towell, J. Shavlik. Training knowledge-based neural networks to recognize genes in dna sequences. *Advances in neural information processing systems*, 3:530–536, 1990.
- [37] M. Pang, K. Ting, P. Zhao, Z. Zhou. Improving deep forest by confidence screening. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1194–1199. IEEE, 2018.
- [38] S. Ni, H. Kao. Psforest: Improving deep forest via feature pooling and error screening. In *Asian Conference on Machine Learning*, pages 769–781. PMLR, 2020.
- [39] G. Hinton, O. Vinyals, J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- [40] L. Utkin, M. Kovalev, A. Meldo. A deep forest classifier with weights of class probability distribution subsets. *Knowledge-Based Systems*, 173:15–27, 2019.

攻读硕士学位期间的主要研究成果

•

致谢

能在浙大度过我学生生涯的最后一段时光，我感到十分幸运。在校两年半的岁月如白驹过隙，一晃而过，在此期间，我收获了很多，也成长了很多。

首先，非常感谢我的导师张国川老师在过去的两年半时间里对我学业上的悉心栽培和指导、生活上的包容和理解，这些对我意义重大，让我度过了充实而快乐的研究生时光。

此外，我还要感谢实验室所有的同窗，相遇即是缘分，能和各位共同走过这一段路，然后同登彼岸，各奔前程，我感到十分荣幸。