

# 基于机器学习分类算法的晃动手机生成音乐方法

林巧民, 燕城成, 张台国

(南京邮电大学教育科学与技术学院, 江苏南京 210023)

**摘要:** 音乐启蒙教育对于儿童的性格塑造、听力与记忆力锻炼、审美能力提升具有重要作用。为此,设计一种易于儿童学习音乐的教学辅助方法。通过对智能手机自带传感器数据的处理,将不同角度的晃动方式区分为不同音符,再通过连续晃动达到演奏简谱的效果。选取加速度传感器,用于反映出手机在空间中晃动的方向变化。为提高动作识别的准确度与反馈的实时性,使用机器学习训练传感器数据分类模型并进行参数调优。实验结果表明,模型识别准确率可达95%以上。将移动应用接入分类模型后,通过不断设计迭代,尤其是引入人一机一曲三者协同机制后,晃动实验数据反馈及时且100%准确识别,从而使用户轻松晃动手机即可精确演奏出简谱中的音符序列。

**关键词:** 传感器;动作识别;机器学习;音乐生成

**DOI:** 10.11907/rjtk.231577

开放科学(资源服务)标识码(OSID):



中图分类号: TP181

文献标识码: A

文章编号: 1672-7800(2025)001-0049-08

## A Music Generation Method for Shaking Mobile Phones Based on Machine Learning Classification Algorithms

LIN Qiaomin, YAN Chengcheng, ZHANG Taiguo

(College of Educational Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** Music enlightenment education plays an important role in shaping children's personalities, improving their listening and memory skills, and enhancing their aesthetic abilities. To this end, design a teaching aid method that is easy for children to learn music. By processing the data from the built-in sensors of smartphones, different angles of shaking are distinguished into different notes, and then continuous shaking is used to achieve the effect of playing simplified music. Select an acceleration sensor to reflect the directional changes of the phone's movement in space. To improve the accuracy of action recognition and real-time feedback, machine learning is used to train a sensor data classification model and optimize its parameters. The experimental results show that the model recognition accuracy can reach over 95%. After integrating the mobile application into the classification model, through continuous design iterations, especially the introduction of the human-machine-music collaboration mechanism, the shaking test data feedback is timely and 100% accurate for recognition. Users can easily shake their phones to accurately play the note sequence in the simplified music score.

**Key Words:** sensor; action recognition; machine learning; music generation

### 0 引言

2014年12月6日,南京师范大学许卓娅教授在全国幼

儿园艺术教学交流研讨会上讲:“对于幼儿园的孩子来说,音乐教育不是教授音乐的教育,而是让孩子通过音乐活动得到发展,最终成长为一个完整的人。”音乐教育可显著促进幼儿智力发展,提升其情感表达和理解能力,辅助其进行社交活动,对于提升幼儿的综合素养具有重要作用。随

收稿日期:2023-06-02

扫描二维码阅读全文:



基金项目:南京邮电大学实验室工作研究课题(2023XSG12)

专利编号:国家发明专利(202310088404.X)

作者简介:林巧民(1979-),男,博士,CCF会员,南京邮电大学教育科学与技术学院副教授、硕士生导师,研究方向为物联网、计算机音乐等。

着机器学习技术的发展,智慧教育逐渐延伸到幼儿教育领域。设计开发易于幼儿上手学习音乐的教学辅助软件成为音乐启蒙教育的有效途径之一。

### 1 相关研究

幼儿肢体协调性欠佳,不能完成复杂的软件操作,晃动手机生成音乐是一种简单且有趣的方法。该方法主要涉及机器学习技术,特别是分类算法。通过算法分析手机传感器采集的运动数据,并将其转化为音乐元素,如音符、节奏等,为幼儿提供了一种新颖的音乐创作体验。

分类识别是基于手机传感器的动作识别的核心。不同分类算法对于不同动作的适用性不同,没有哪一种分类算法对于任何人体动作的识别准确率是远高于其他分类算法的,使用何种分类识别算法需要根据实验结果和具体应用场景决定。基于模板匹配的分类识别方法是最早的序列数据分类算法,通过动态规划的方式实现,通过计算待分类序列与模板序列的相似度值并比较得出最小相似度值得到该序列的分类值,实现简单,但准确率并不高<sup>[1-2]</sup>。随后基于统计分析的分类识别方法开始取代模板匹配,常见分类算法包括逻辑回归、k-近邻、决策树、支持向量机(Support Vector Machine, SVM)和贝叶斯等。这些方法可以通过有限的样本构建出鲁棒性较强的分类模型,实用性较好<sup>[3]</sup>。

用于人体动作手势识别的手机传感器主要包括加速度传感器、磁力传感器和陀螺仪传感器,其中加速度传感器又分为带重力和不带重力两种<sup>[4-5]</sup>。采集到的传感器数据需要进行平滑去噪和数据切割处理。平滑去噪是指降低采集数据过程中出现的噪声,减少对于后续步骤的影响,如可能导致的模型过拟合问题<sup>[6]</sup>。通常采用滤波器进行传感器信号去噪,如移动均值滤波器、高斯滤波器、巴特沃斯滤波器和3阶滑动平均滤波器<sup>[7]</sup>。数据切割是指将整段数据分割成代表某一特征的一段段数据。将较长的数据流切割为长度相似的数据样本有利于后续特征提取与分类识别。滑动窗口法是目前最常用的数据切割方法,需要设定窗口大小和覆盖率,滑动窗口过短或过长均会导致识别准确率降低<sup>[8]</sup>。窗口大小是不固定的,需要根据不同人体活动进行多次实验得出,不同识别方法的窗口大小也有区别。时域特征可反映传感器信号的分布,直接针对传感器时序数据提取特征的方法简单、计算量小,比较适合对实时性要求较高的场景<sup>[9]</sup>。与时域特征观察信号变化幅度不同,频域特征通过观察功率谱来提取特征。进行频域特征提取时需要将信号从时域转化到频域,因此计算量很大。将时域特征转化到频域的方法主要包括快速傅里叶变换(Fast Fourier Transform, FFT)和离散余弦变换(Discrete Cosine Transform, DCT),以往研究大多采用FFT的方式<sup>[10]</sup>。频域特征大多是参考语音特征和行为特征筛选出

来的,结合时域特征可作为预备的原始特征集合。

为了提高动作识别的准确率,获得及时良好的交互反馈,本文结合机器学习的分类算法提出一种基于训练模型与传感器动作数据进行音乐自动生成的方法。本文工作不仅仅是聚焦于传感器数据的处理以及模型的训练调优,还设计算法将模型运用到移动应用中,是一种新颖的交互式音乐生成方式。

### 2 系统模块设计

图1左为以智能手机为中心的空间坐标系示意图,右为手机晃动方向与音阶的平面对应关系。数字对应7个音高,当想要演奏某个音符时,只需朝对应方向晃动,手机传感器在检测到晃动趋势后,程序播放音符,然后将设备归位到初始的中间状态,如此往复演奏完整歌谱。后续的模型训练和数据切分均基于上述晃动规则。系统分为模型训练、数据切分和音符播放3个模块,其中模型训练模块在服务端部署,数据切分和音符播放模块在手机上运行。

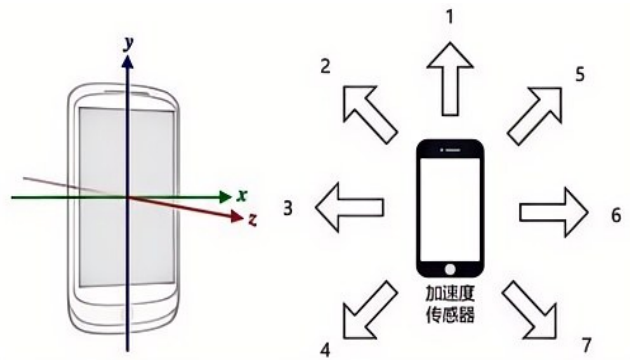


Fig. 1 Space coordinates and musical scale distribution

图1 空间坐标系与音阶分布

#### 2.1 模型训练模块

图2为模型训练模块设计。该模块主要负责对传感器数据分类模型进行训练,包括数据采集、数据集构建、特征提取、分类算法选择和参数调优等步骤。首先,采集加速度传感器接收的数据,依据晃动动作与对应音符规则构建数据集;其次,提取时域特征,通过实验比较构建最优特征集;最后,选择SVM分类算法,并进行参数优化,选择最优数据序列识别长度作为窗口大小。

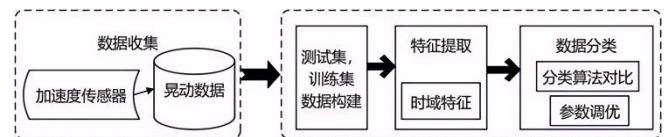


Fig. 2 Model training module design

图2 模型训练模块设计

#### 2.2 数据切分模块

虽然智能手机对于传感器数据流的检测已经成熟,可精确检测手机加速度的变化,但是仍然无法准确判断动作

的开始点。提前或延迟接收一条数据会导致被预测的音符数据序列存在偏移,从而影响模型预测准确度。因此,本文设计了基于程序控制晃动节奏的数据切分算法,控制固定节奏算法输入为演奏时长  $m$ ,单个晃动动作时长  $n$ ,回调触发间隔时长  $t$ ,接收传感器数据时间点  $startTime$ ,停止接收时间点  $endTime$ ,屏幕颜色变蓝时间点  $screenColorBlueTime$ ,屏幕颜色变白时间点  $screenColorWhiteTime$ ;输出为无。数据切分算法伪代码如下:

#### 算法1 数据切分算法

输入:原始传感器数据序列  $D$ ,传感器数据预测序列长度  $p$ ,用于预测的传感器数据片段容器  $E$

输出:无

```

1 实时监听传感器数据事件 event:
2   if 传感器状态 == 可接收数据:
3       if E.length == p:
4           调用模型
5       else E.add(event)
6   return

```

控制节奏算法设置定时器,通过固定的时间间隔回调需要执行的代码段,参数  $m$  即用户演奏某个歌谱的时长,参数  $n$  即自定义的单个动作晃动时长,那么  $m$  与  $n$  的关系即为  $m=n \times$  需要晃动的音符数量。通过计算  $m \% n$  便可以得到某个晃动周期固定的值序列,而  $startTime$ 、 $endTime$ 、 $screenColorBlueTime$ 、 $screenColorWhiteTime$  则需根据具体实验进行设置,其与传感器性能、手机性能密切相关。系统除了设置  $startTime$ 、 $endTime$  参数更改传感器状态全局变量以供算法 1 进行数据切分的判断外,还设置了  $screenColorBlueTime$ 、 $screenColorWhiteTime$  参数提醒用户最佳的晃动时间点,白色为默认页面颜色,屏幕变为蓝色即为晃动时间点的提示。控制节奏算法通过输入参数的调整实现了晃动节奏的判断以及不同类型、型号手机的适配。

#### 2.3 音符播放模块

音符播放模块由辅助歌谱以及音乐合成器组成。根据需求分析,该模块需要实现音符的音长和音高两个特性。然而,本文设计的晃动规则只有 7 个方向的区分,只能实现 7 个音阶。为了实现其他音阶以及音长,通过迭代设计,辅助歌谱是目前实际可行、效果较好的方案。通过用户主动选择要演奏的歌谱,程序自动匹配晃动的音符并对应于各个 8 度的相应音阶以及音长。

### 3 实验方法与结果分析

#### 3.1 实验工具

模型训练模块使用到的工具为 Python、Anaconda(用于多环境的控制)、scikit-learn(机器学习相关库)、pandas、matplotlib、sklearn\_porter(基于 sklearn 实现的模型格式化开源项目)。

#### 3.2 预先考量

使用机器学习分类模型而不是直接分析晃动数据、设计简单算法来实现音符识别,是由于直接设置每个动作的传感器数据阈值,在达到阈值时播放对应音符的效果一般,且通常需要较大力气进行晃动才能正确识别。这是由于后者无法利用整段传感器数据,算法获取的信息量不够。动态时间规整算法可以利用整段序列数据,算法 2 简单地展示了其实现过程。

#### 算法2 改进的动态时间规整算法

输入:模板动作数据序列  $M$ ,测试动作数据序列  $T$ ,预测片段长度  $length$

输出:序列间最短距离  $dist$

```

1 初始化距离 dist,权重二维数组 weight
2 for i=0; i<length; i++:
3   for j=0; j<length; j++:
4     if i<j+length/2 && i>j-length/2:
5       weight[i][j] = M 与 T 的对应序列点的欧式距离
6     else weight[i][j] = -1 //即距离不可达
7     dist = dp(weight) //动态规划计算最短距离
8   return dist

```

代码第 4 行是通过  $i, j$ , 即  $M$  与  $T$  间对应点的关系进行计算量上的优化,通过计算 7 个模板动作数据序列与测试动作数据序列得到最小距离,最小距离对应的模板动作即为匹配得到的对应动作值。共 100 组动作序列,将第 40 组作为模板动作,7 种不同动作匹配精准度如表 1 所示。

Table 1 Matching accuracy for 7 different action

表 1 7 种不同动作匹配精准度

| 分类   | 动作 1 | 动作 2 | 动作 3 | 动作 4 | 动作 5 | 动作 6 | 动作 7 | 匹配精准度/% |
|------|------|------|------|------|------|------|------|---------|
| 动作 1 | 94   | 1    | 0    | 0    | 3    | 1    | 0    | 95      |
| 动作 2 | 14   | 77   | 7    | 0    | 0    | 1    | 0    | 78      |
| 动作 3 | 4    | 2    | 86   | 2    | 0    | 4    | 1    | 87      |
| 动作 4 | 8    | 1    | 6    | 68   | 0    | 3    | 13   | 69      |
| 动作 5 | 24   | 0    | 0    | 1    | 65   | 9    | 0    | 66      |
| 动作 6 | 2    | 0    | 2    | 2    | 3    | 89   | 1    | 90      |
| 动作 7 | 3    | 0    | 2    | 2    | 0    | 12   | 80   | 81      |

第 1 次优化措施应从数据中提取更多信息,如提取序列的最值,将原来的仅用序列最短距离改为  $dist+p*d(M1, T1)+q*d(M2, T2)$ ,其中  $M1, T1$  分别表示模板序列和测试序列 3 个轴的最大值,  $M2, T2$  分别表示模板序列和测试序列 3 个轴的最小值,  $d$  表示欧式距离,  $p, q$  为通过动态计算得出的合理参数,用于提升匹配精确度。经过计算,当  $p=-4, q=5$  时,平均匹配精准度最高。结果见表 2。

第 2 次优化措施为进一步提取序列数据的方差,再加入新的参数进行计算调优,这与模型的训练过程类似。提取数据序列的特征,手动设计特征对应的参数并计算优化参数,利用机器学习相关工具实现实验数据的分类。

#### 3.3 数据采集

数据储存在服务端,采用 NodeJS 的 Express 框架,易于实现简单数据的接口请求。将手机与 PC 设置在同一局域网

Table 2 Matching accuracy for 7 different actions after initial optimization

表2 首次优化后的7种不同动作匹配精准度

| 分类  | 动作1 | 动作2 | 动作3 | 动作4 | 动作5 | 动作6 | 动作7 | 匹配精准度/% |
|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| 动作1 | 90  | 1   | 1   | 6   | 0   | 0   | 0   | 90      |
| 动作2 | 12  | 79  | 7   | 0   | 0   | 0   | 1   | 80      |
| 动作3 | 0   | 0   | 95  | 2   | 0   | 1   | 1   | 96      |
| 动作4 | 3   | 1   | 4   | 84  | 1   | 0   | 6   | 85      |
| 动作5 | 16  | 0   | 0   | 2   | 77  | 4   | 0   | 78      |
| 动作6 | 2   | 0   | 3   | 2   | 5   | 86  | 1   | 87      |
| 动作7 | 3   | 1   | 1   | 0   | 0   | 4   | 90  | 90      |

网内, App中设置PC服务端ipv4地址,  $x$ 、 $y$ 、 $z$ 分别实时记录不带重力的加速度传感器三轴的动态变化值。App中设置3个按钮, 分别为开始发送数据、停止发送数据和测试发送数据。PC服务端接收每一次动作传输的数条传感器数据并保存到excel表格中, 命名为对应动作值1-7。实验共采集100组动作数据, 每一组为7个晃动动作, 即数据集大小为700, 其中80%作为训练数据集, 20%作为测试数据集。

### 3.4 数据预处理

数据预处理主要考虑降噪和窗口大小的选择。如表3所示, 动作1的数据序列长度为10。由图3可见, 动作1的三轴值曲线变化明显, 完全能够体现动作的变化趋势, 所以综合考虑, 实验舍去了数据降噪的步骤。关于窗口大小的选择, 实验综合比较图3、图4和图5, 发现不同动作展现出的动作趋势(即横坐标从0开始的序列片段)有明显差异。例如动作1是向上挥动, 那么 $y$ 轴加速度数值会明显增大,  $x$ 轴数据仍在0附近; 动作2是向左上角挥动,  $y$ 轴加速度数值变化与动作1相同, 但 $x$ 轴数值会明显朝负方向增大, 图中的变化与实验分析是一致的。具体要提取的数据序列片段长度需要进一步实验验证, 这里先以函数的形式展示对原始数据的处理, 将所需验证的最佳长度以参数的形式表示, 在后续分类算法比较中论证最佳的参数长度。

Table 3 Data sequence for action 1 of group 11

表3 第11组动作1的数据序列

| 序号 | $x$        | $y$        | $z$        |
|----|------------|------------|------------|
| 0  | -0.084 534 | 0.053 020  | 0.027 538  |
| 1  | -0.051 775 | -0.032 528 | 0.023 780  |
| 2  | 0.075 795  | 0.338 819  | -0.201 583 |
| 3  | 0.152 003  | 2.215 143  | 1.606 664  |
| 4  | -0.906 775 | -0.589 713 | 0.602 139  |
| 5  | 0.230 777  | -2.896 847 | 1.651 848  |
| 6  | 0.192 059  | -1.908 278 | -1.067 325 |
| 7  | 0.167 545  | -0.230 824 | -0.060 375 |
| 8  | -0.518 895 | 0.854 473  | -0.960 262 |
| 9  | 0.184 169  | 1.064 102  | -0.264 194 |
| 10 | 0.188 785  | 0.448 473  | 0.582 900  |

### 3.5 特征选择

特征选择直接影响模型识别的准确率, 观察模型与选取特征的相关性并进行多次实验, 选择有价值的特征构建

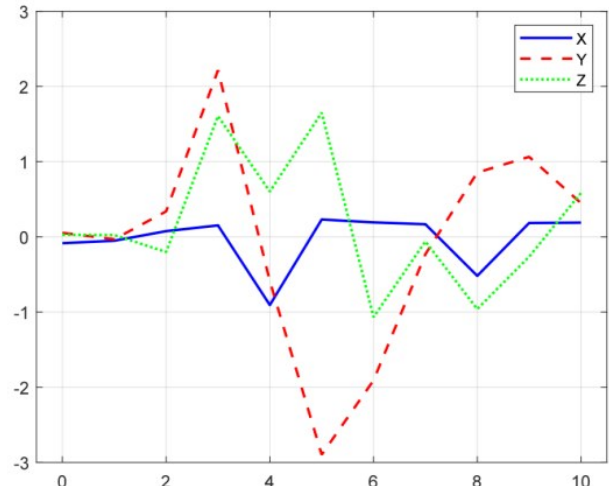


Fig. 3 Line chart for action 1 of group 11

图3 第11组动作1的折线图

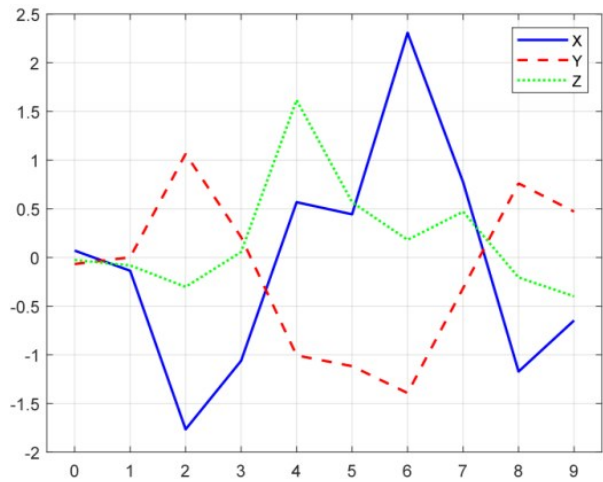


Fig. 4 Line chart for action 2 of group 11

图4 第11组动作2的折线图

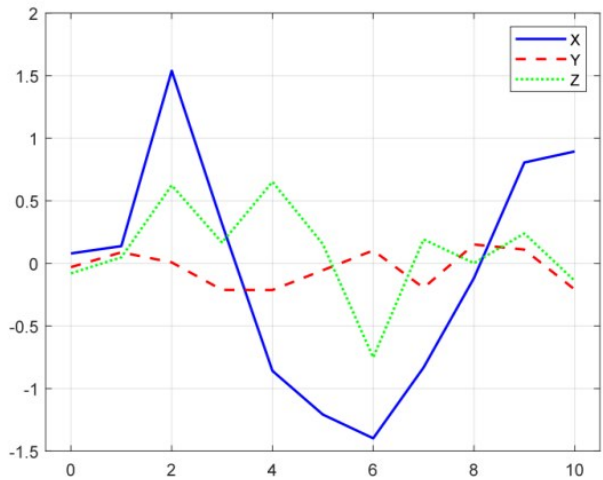


Fig. 5 Line chart for action 6 of group 11

图5 第11组动作6的折线图

特征集。本文设置了4组特征集: 第一组为3轴的平均值和方差; 第二组为3轴的平均值、方差、最大值以及最小

值;第三组为3轴的平均值和方差,x轴、y轴的最大值、最小值;第四组为x轴、y轴的平均值、方差、最大值以及最小值。以SVM为例,以核函数参数选取高斯核函数, $\gamma$ 值设置为0.2,单个动作数据序列片段长度设为5,使用sklearn的学习曲线learning\_curve函数,采用该函数而不是默认的svc函数直接得出结果的原因是学习曲线训练时引入了交叉验证集,使真正的测试集结果准确性更有参考价值。同时,learning\_curve函数会对同一数量的训练样本进行多次计算,可以通过观察平均值和方差得到更加稳定准确的结果。

4组特征集的训练结果如图6、图7、图8和图9所示。图中横轴为训练的样本数量,通过设置learning\_curve函数的train\_sizes参数,例如设置为(0.1,1.0,5),表示将训练样本从0.1~1分为5等份;同时设置test\_size=0.2,那么训练样本总共为560,分成5份即为56、182、308、434、560,对应点的横轴坐标与实验计算的分配比例是吻合的。竖轴为每次训练的得分。可以看出,在特征集最少的第一组中,训练分数与测试分数均趋于同一水平,收敛在0.95左右;在第二组增加最值的情况下,也是特征集最丰富的情况下,模型在训练样本增加的过程中逐渐趋于稳定,收敛在0.975左右,可以看出最值特征具有一定价值;在第三组中去除了可能对模型影响不大的z轴最值,结果与第二组相差不大,也收敛在0.975左右;第四组进一步移除z轴的其他特征,可以看到收敛分数相较第二组、第三组明显下降,但仍高于第一组,收敛在0.95~0.975之间。以上实验结果说明,x轴、y轴的最值特征具有一定价值,z轴的平均值和方差特征具有一定价值。综合考虑,选择第三组作为最终模型的特征集。

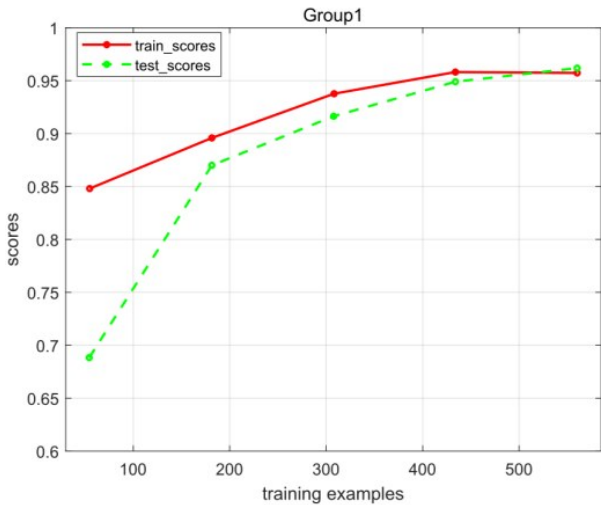


Fig. 6 The first group  
图6 第一组

### 3.6 分类算法调优

用于分类问题的算法有很多,本文选择适合进行传感器数据分类的k-近邻、逻辑回归和SVM算法。3种算法的前置数据预处理和特征值选择均相同,训练结果如图10、

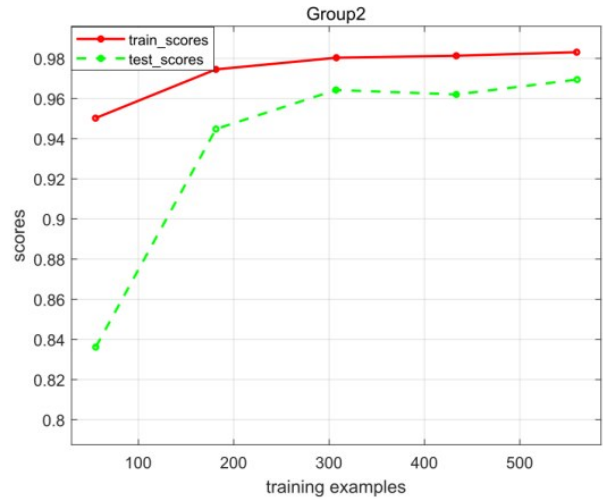


Fig. 7 The second group  
图7 第二组

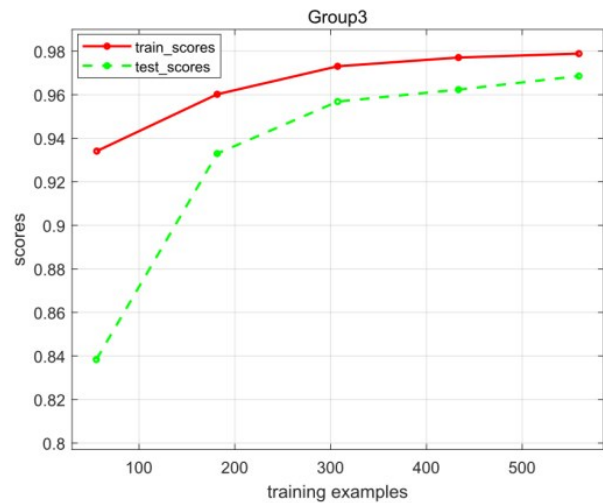


Fig. 8 The third group  
图8 第三组

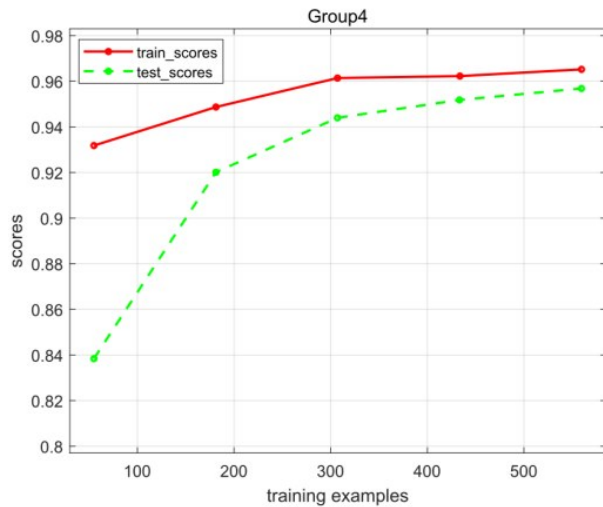


Fig. 9 The fourth group  
图9 第四组

图11和图12所示。k-近邻算法设置n\_neighbors参数为3,当训练样本数量为560时,测试分数平均值在0.96左右,

但上下浮动范围较大,当训练样本增大时也不会有明显的收敛迹象;逻辑回归算法添加 pipeline 增加多项式参数,设置二阶多项式,penalty 范式选择 L1,可以看出稍微进行参数调节算法准确率就有很大提升;SVM的核函数为高斯核函数, $\gamma$  值设置为 0.2,综合训练分数和交叉测试分数均高于其他两种算法。

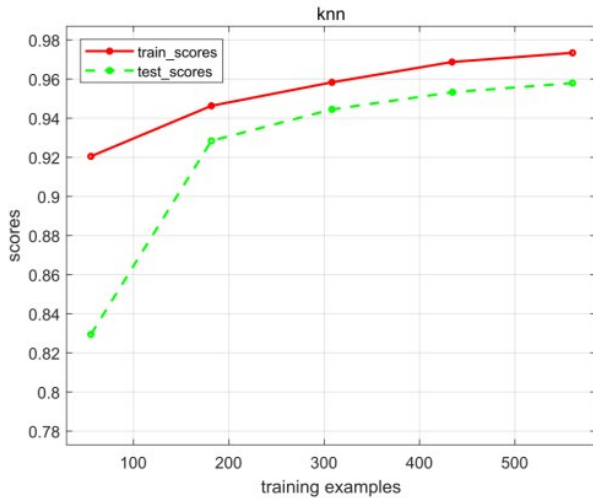


Fig. 10 K-nearest neighbor  
图 10 k-近邻

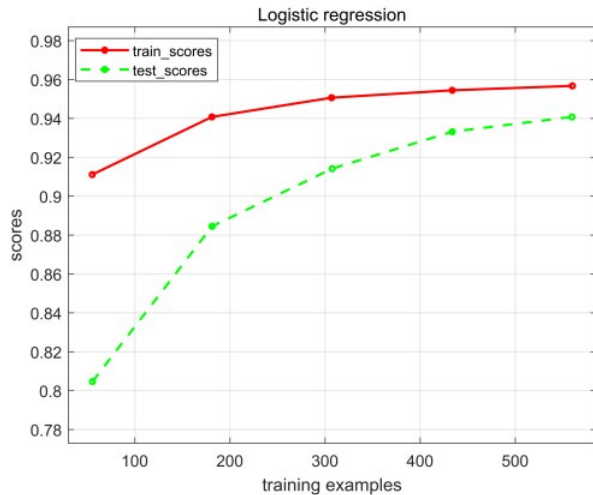


Fig. 11 Logistic regression  
图 11 逻辑回归

需要注意的是,本文并未对逻辑回归算法进行更多参数测试和调整,因为实验使用的 sklearn-porter 库不支持逻辑回归模型转化,最初的方案也并不是使用 sklearn-porter 库而是直接通过手机网络调用服务端程序,再通过服务端程序调用模型进行预测后将结果返回给手机。然而,该方案消耗的网络时长约为 800 ms,在实际测试中,假设一个晃动动作为 2 s,在做出某个方向动作的同时程序也检测到完整的动作数据序列片段,此时已过去约 900 ms,再消耗 800 ms 做预测,在返回预测音符后,此时已过去 1.7 s,留给设备播放音符的时长就只剩下不到 300 ms,这显然不合理。因此,本文选择使用第三方库转化的方式使设备可以

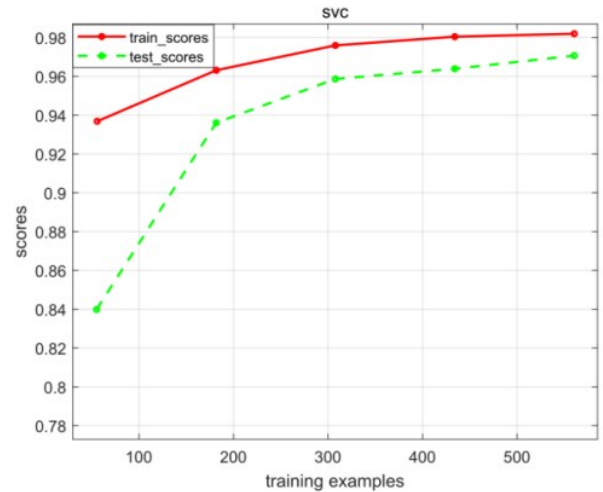


Fig. 12 Support vector machine  
图 12 SVM

直接调用模型进行预测,相较原有方案至少省去了 700 ms 的网络消耗,而留给演奏播放音符的时间也扩展到 1 s。

sklearn-porter 可支持 SVM 模型转化,因为本文将重点放在 SVM 模型调优上<sup>[11]</sup>。如图 13 所示,当核函数为多项式且项数为 5 时,SVM 模型的训练分数和测试分数均最终收敛在 0.77 左右,说明多项式核函数无法满足所设置的特征集,欠拟合现象明显。

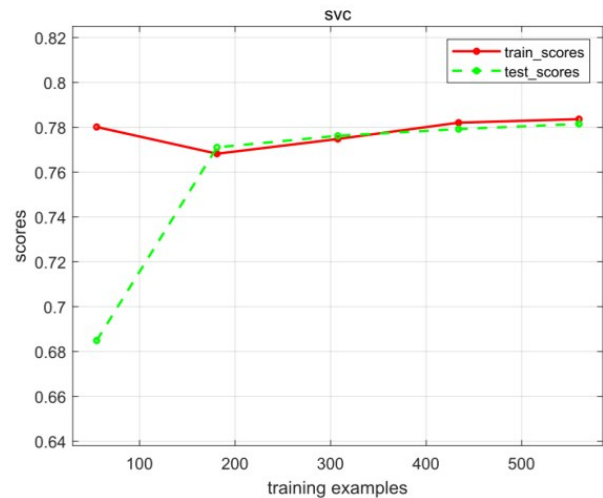


Fig. 13 Polynomial kernel function  
图 13 多项式核函数

高斯核函数可将输入的特征映射到无限多维,但同样需要考虑  $\gamma$  值的设置,否则可能会出现过拟合现象<sup>[12]</sup>。本文采用 sklearn 的模型参数工具函数 GridSearchCV,其中的关键参数为 param\_grid,该参数除了设置模型的选择项参数外(例如选择核函数为线性还是多项式),在数值参数处可以指定数组,设置为 np.linspace(0, 2, 40),表示从 0~2 切分为 40 份,即为  $\gamma$  值。GridSearchCV 类似于学习曲线,也会在同一训练样本数量、参数值下进行多次运算,从而通过获取均值、方差的形式展现出训练分数和测试分数的变化趋势以及收敛情况,且其支持交叉测试。如

图 14 所示, 横轴表示动作数据序列长度, 竖轴为对应动作数据序列长度下最佳  $\gamma$  值时的模型交叉测试分数。动作数据序列长度选取 3~7 的区间, 即以识别动作趋势的方式训练模型, 没必要将更长的动作数据序列纳入实验范畴。可以看出, 当动作数据序列长度为 5 时, 模型分数远高于其他情况。

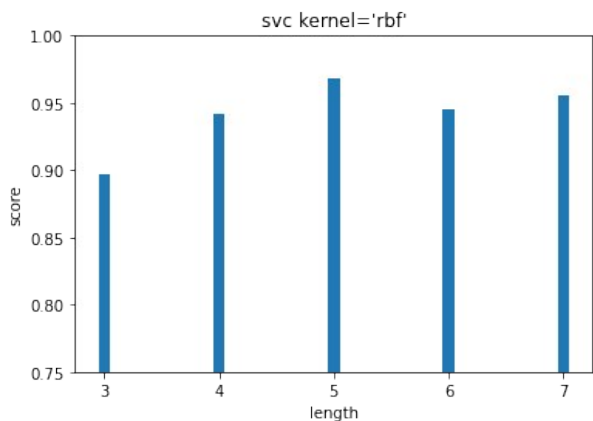


Fig. 14 Gaussian kernel

图 14 高斯核函数

图 15 为动作数据序列长度为 5 时  $\gamma$  参数与模型评分的关系。可以看出, 当  $\gamma$  值为 0.2 时, 模型训练分数与测试分数较接近且处于较高水平, 但随着  $\gamma$  值的增大, 训练分数慢慢提高最终接近 1, 测试分数则逐渐下降, 两者的差距越来越大, 这是由于模型过拟合现象越来越明显。当  $\gamma$  值为 0.205 128 时, 模型评分最高, 为 0.967 857。

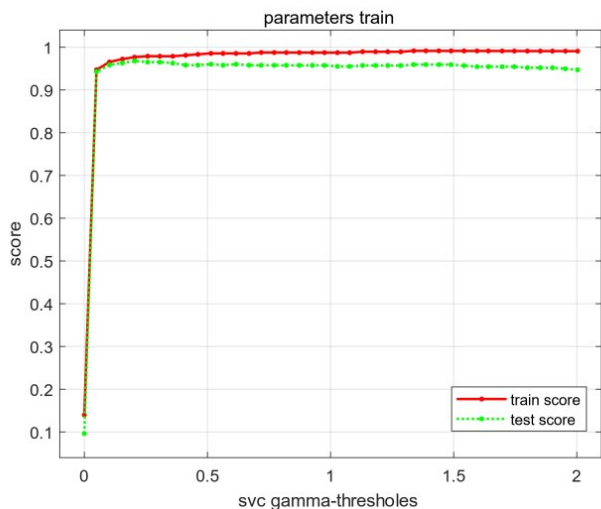


Fig. 15  $\gamma$  parameters and model rating

图 15  $\gamma$  参数与模型评分

采用 sklearn 的 classification\_report 可以查看模型每一类的查准率、召回率和 F1-score。表 4 为 SVM 模型的输出结果, 共包含 0~6 的 7 个动作标签。除标签 3 外, 其余标签的查准率均达到 0.95, 标签 1、5、6 的召回率甚至达到 1.0。从 F1-score 来看, 仅标签 3 未达到 0.95, 其余标签均在 0.95 以上。这是由于标签 3 对应于晃动规则的动作 4, 即向左下角挥动, 由于实验时是右手挥动, 向左下角挥动的幅度

会小一些, 可能会有一定的概率被误判为其他动作。而标签 1 和标签 5 分别对应晃动规则中的动作 2 和 6, 即水平向左和水平向右, 人体比较容易且自然做到, 识别准确率也更高, 因此 F1-score 均为 1.0。

Table 4 SVM model output results

表 4 SVM 模型输出结果

| 标签  | 查准率  | 召回率  | F1-score |
|-----|------|------|----------|
| 0   | 0.96 | 0.96 | 0.96     |
| 1   | 1.00 | 1.00 | 1.00     |
| 2   | 1.00 | 0.95 | 0.97     |
| 3   | 0.86 | 0.90 | 0.88     |
| 4   | 1.00 | 0.94 | 0.97     |
| 5   | 1.00 | 1.00 | 1.00     |
| 6   | 0.96 | 1.00 | 0.98     |
| 准确率 |      |      | 0.96     |

### 3.7 基于节奏的程序控制

采用“人—机—曲”三者协同机制, 即基于人演奏曲子的固有节奏进行程序的传感器状态转换控制, 演奏曲谱时只需根据屏幕变蓝的提示做出所需音符对应的晃动动作即可。测试设备型号为 Android RMX2202, 经过多次调试实验, 设置 startTime 为 0.0, screenColorBlueTime 为 1.7, screenColorWhiteTime 为 0.3, endTime 为 0.1, 单个晃动动作时长  $n$  为 2 s, 演奏时长为 100 s, 每个动作的晃动次数为 10 次, 此时, 所有动作识别准确率均达到 100%。由于是程序控制节奏, 人机配合十分自然流畅。

### 3.8 音符播放

实验中使用的音乐合成器软件基于 javax.sound.midi-for-Android 开源库实现。由于使用 Dart 作为开发语言, 需要编写 Flutter 插件进行不同语言调用的 api 转化。根据需求封装了 3 个 api, 分别为 prepare、playMidiNote 和 stopMidiNote, 其中 prepare 方法用于加载 .sf2 格式的模拟钢琴声音, playMidiNote 和 stopMidiNote 则是调用 midi 播放音符和停止播放, 支持音阶和音高两个参数。关于辅助歌谱功能, 用户可事先选择想要演奏的曲谱, 开始后只需根据辅助歌谱的提示晃动即可, 程序会实时进行晃动音符与对应音阶及音长的自动匹配。

## 4 结语

儿童具有爱玩的天性, 基于肢体动作的音乐教学可更加吸引其注意力。本文结合机器学习分类算法 SVM 实现了一种基于智能手机的具身交互式音乐演奏方式, 达到了较高的动作识别准确率和及时良好的反馈效果。未来可基于轻量级深度学习技术, 如 MobileNet<sup>[13]</sup>、SqueezeNet<sup>[14]</sup> 等进行新的尝试, 这类模型可以在较小的数据集上实现良好性能<sup>[15]</sup>, 有助于设计更新颖的交互方式, 同时实现更广泛范围的音阶映射。

## 参考文献:

- [1] KURUPPU G, MANOJ C, KODITUWAKKU S R, et al. Comparison of different template matching algorithms in high speed sports motion tracking [C]//2013 IEEE 8th International Conference on Industrial and Information Systems, 2013:445-448.
- [2] MARGARITO J, HELAOUI R, BIANCHI A M, et al. User-independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach[J]. IEEE Transactions on Biomedical Engineering, 2016, 63(4):788-796.
- [3] ZHANG J W, JIA H Y, ZHANG N. Alternate support vector machine decision trees for power systems rule extractions [J]. IEEE Transactions on Power Systems, 2023, 38(1):980-983.
- [4] LEMOYNE R, MASTROIANNI T, MCCANDLESS C, et al. Implementation of a smartphone as a wearable and wireless accelerometer and gyroscope platform for ascertaining deep brain stimulation treatment efficacy of Parkinson's disease through machine learning classification [J]. Parkinson, 2018, 7(2):19-30.
- [5] JAVED A R, FAHEEM R, ASIM M, et al. A smartphone sensors-based personalized human activity recognition system for sustainable smart cities [J]. Sustainable Cities and Society, 2021(71):1-12.
- [6] YANG X H, LI X J, GUAN Y, et al. Overfitting reduction of pose estimation for deep learning visual odometry [J]. China Communications, 2020, 17(6):196-210.
- [7] ALI H, YAQUB R. A low complexity linear moving average filtering technique for PAR reduction in OFDM systems [J]. International Journal of Communication, Network and Systems, 2018, 11(10):199-215.
- [8] LI Q H, ZHANG Z, FENG C, et al. Human motion prediction using optimized sliding window polynomial fitting and recursive least squares [J]. The Journal of China Universities of Posts and Telecommunications, 2021, 28(3):76-85.
- [9] NAM Y, PARK J W. Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor [J]. IEEE Journal of Biomedical and Health Informatics, 2013, 17(2):420-426.
- [10] LAI L Q, CHEN Y, CHEN K T, et al. Reconstruction and fitting of second-harmonic signals by wavelength modulation spectroscopy method based on fast Fourier transform [J]. Chinese Optics Letters, 2022, 20(9):68-72.
- [11] Scikit-learn. Support vector machines — scikit-learn [EB/OL]. <https://scikit-learn.org/stable/modules/svm.html>.
- [12] WANG D. Research on support vector machine clustering algorithm based on weighted multi width gaussian kernel function [D]. Qingdao: Ocean University of China, 2011.  
王栋. 基于加权多宽度高斯核函数的支持向量机聚类算法研究 [D]. 青岛: 中国海洋大学, 2011.
- [13] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:4824-4832.
- [14] FORREST N I, HAN S, MATTHEW W M, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size [DB/OL]. <https://arxiv.org/abs/1602.07360>.
- [15] PENG L. Piano players' intonation and training using deep learning and MobileNet architecture [J]. Mobile Networks and Applications, 2023, 28(6):2182-2190.

(责任编辑:尹晨茹)