

上海工程技术大学

第十四届研究生数学建模竞赛

学 院	机械与汽车工程学院
--------	-----------

专 业	机械工程
--------	------

班 级	
--------	--

参赛队号	
------	--

队员姓名	1.
------	----

2.

3.

上海工程技术大学

第十四届研究生数学建模竞赛

题 目 基于机器学习的生命体征预测模型

摘 要：

本文研究的是利用已知电子信号建立数学模型从而研究生命体征的问题。

问题一中，我们对已知的电子信号进行数据处理并进行聚类分析。首先，我们对呼吸、体动、心率这三者进行皮尔逊部分相关性分析，并发现心率和体动具有较大的相关性，与实际相符。接着我们对“opticalpower”数组进行分析，频谱分析并通过拟合概率密度函数得到了信号的特征分布，并得出结论光纤电信号预测呼吸和心率的模型是相似的。

问题二中，我们应用机器学习的办法，对电子信号估计生命体征模型进行建立，对于每个生命体征，我们使用了 4 种模型进行建立；其中两个生命体征的表现良好，可以直接应用。对于体动值这一生命体征，我们提出了类比神经网络结构的串联模型，并加以评估，发现可以较好的预测体动值。

问题三中，我们查阅相关资料，决定对不宁腿综合征这一疾病进行分析。分析疾病特点并结合问题二中的体动特征图，对目标的体动这一特征进行分析，针对表现异常的电子信号序列，单独分析，并得出体动值和不宁腿综合征的联系。

问题四中，我们结合文献介绍了一些需要被关注的生命体征以及部分的监测方法。

关键词：机器学习 生命体征 随机森林 聚类分析 频谱分析

第一章 问题重述

生命体征监测垫是利用光纤传感器人睡觉时的静息状况进行的准确科学的检测。由于传感器获得的是离散时间的电子信号序列，分析这些序列的特征以及这些序列反映的生命体征特征是我们面对的一个实际课题。

本文主要就某医疗器械企业利用其生命体征监测垫收集的志愿者睡眠阶段的数据集，建立了合理的数学模型并分析研究下面的问题：

(1) 利用 **K-means** 聚类对现有数据集中的电子信号序列算法进行聚类分析，并对每一个聚类，分析其电子信号序列的特征及相应生命体征的特征。

(2) 电子信号序列是一个 50×1 的数组，而其余 3 个数据均是单个的数，而机器学习对于输入和输出的规则要求不高，因此我们使用机器学习的办法来建立通过电子信号估计生命体征数据的数学模型。

(3) 查阅资料了解不宁腿综合征的生命体征，并分析其电子信号序列的特征。

(4) 除了呼吸频率、心率、体动这些生命特征外，查阅资料并分析了体温、血压等生命体征等也需要受到关注。

第二章 数据处理

2.1 皮尔逊部分相关性分析

首先对所有 json 文件中的 4 个元素进行提取，进行相关性分析。

皮尔逊相关系数 (Pearson correlation coefficient) 是一种用于度量两个变量之间线性相关程度的统计量。它衡量了两个变量之间的线性关系的强度和方向。这个系数的值介于-1 和 1 之间，其中-1 表示完全的负相关，1 表示完全的正相关，0 表示没有线性相关。

数学上，皮尔逊相关系数可以通过以下公式计算：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

其中，r表示皮尔逊相关系数， X_i 和 Y_i 分别表示第i个数据点的两个变量的值， \bar{X} 和 \bar{Y} 分别表示X和Y的平均值。

通过计算皮尔逊相关系数，我们可以了解两个变量之间的线性关系。如果r的值接近 1 或-1，表示两个变量之间存在强烈的线性相关性，正值表示正相关，负值表示负相关。如果r的值接近 0，则表示两个变量之间几乎没有线性关系。

我们发现原始数据中，除” opticalpower” 的数据是 50x1 的矩阵，其余的” breath” ,” heart_rate” ,” totalMotion” 均为单个数字，所以先进行这三者相关性的分析。

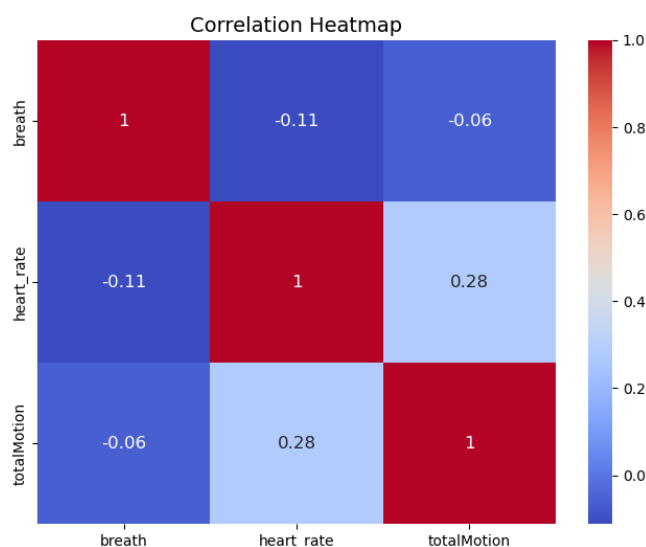


图 2-1 三元素相关系数矩阵热力图

根据上面的热力图进行分析，发现这三者中，心率(heart_rate)和体动值(totalMotion)的相关系数最大。随着身体运动的剧烈，心率会升高^[1]，该相关系数正能说明此结论。

2.2 聚类分析

而对于另外的一组”opticalpower”的处理，Shen^[2]等人提出了支持向量机(SVM)结合 K-means 聚类算法的方法，对 100000 个心电样本进行分类。所以以此为基础，我们打算先对”opticalpower”进行 K-means 聚类分析。

K-means 聚类算法是一种常用的无监督学习算法，用于将一组数据点分成具有相似特征的不同群集（簇）。它通过迭代的方式将数据点分配给最近的簇，并更新簇的中心点，直到达到收敛状态。

数学上，K-means 聚类算法可以表示为以下步骤：

（1）初始化：选择要分成的簇的数量 K，并随机选择 K 个数据点作为初始的簇中心。令簇中心为 C_1, C_2, \dots, C_K

（2）分配数据点：对于每个数据点 X_i ，计算它与各个簇中心的距离，并将其分配给距离最近的簇。令 $d(X_i, C_j)$ 表示数据点 X_i 到簇中心 C_j 的距离。则数据点 X_i 属于第 j 个簇的概率为：

$$P(C_j|X_i) = \begin{cases} 1 & \text{if } j = \arg \min_k d(X_i, C_k) \\ 0 & \text{otherwise} \end{cases}$$

（3）更新簇中心：对于每个簇 C_j ，计算其中所有数据点的平均值，将该平均值作为新的簇中心：

$$C_j = \frac{\sum_{X_i \in C_j} X_i}{|C_j|}$$

其中 $|C_j|$ 表示簇 C_j 中的数据点数量。

（4）重复分配数据点和更新簇中心，直到簇中心不再改变或达到预定的迭代次数。K-means 聚类算法通过迭代的方式不断优化簇中心和数据点的分配，使得同一簇内的数据点更加相似，不同簇之间的数据点更加不相似。最终得到的簇划分可以帮助我们发现数据的内在结构和模式。

首先使用 K-means 聚类算法对光纤反射转换的电信号（下称光纤维电信号）进行聚类，初始选择的簇数是 5，但经过分析发现有两簇数据的方差近似，因此最终选择合并为 4 个簇，其中心点如下图 2-2 所示。

```

Cluster 0 centroid: [-18.85634807 -18.85827221 -18.85828345 -18.85742583 -18.85826269
-18.85805986 -18.85928683 -18.85950913 -18.85944512 -18.85945809
-18.85953248 -18.85962849 -18.85970029 -18.85920465 -18.85918346
-18.85848197 -18.85806072 -18.85807889 -18.85788859 -18.85782545
-18.85687873 -18.85614869 -18.85590606 -18.85592985 -18.8562806
-18.85652106 -18.8567758 -18.85678531 -18.85698772 -18.85680088
-18.85713736 -18.85736441 -18.85786048 -18.8584106 -18.85888937
-18.85803304 -18.85792146 -18.85838639 -18.85826399 -18.85833232
-18.8586182 -18.8587047 -18.85910475 -18.85896203 -18.85923536
-18.86002898 -18.8601198 -18.86013018 -18.86008996 -18.86013018]
Cluster 1 centroid: [-17.80556306 -17.80880358 -17.80965068 -17.80770188 -17.80768731
-17.80809609 -17.81003131 -17.81044639 -17.81038062 -17.81047868
-17.81012602 -17.81042828 -17.81048302 -17.81075377 -17.81003525
-17.80975406 -17.80920902 -17.80940396 -17.80898612 -17.80885222
-17.80886403 -17.80812484 -17.80797243 -17.80734449 -17.80728877
-17.80761288 -17.80805848 -17.80780506 -17.80834991 -17.8085222
-17.8085415 -17.80885261 -17.80903357 -17.80954199 -17.80755381
-17.80675002 -17.8068727 -17.80861396 -17.8085411 -17.80911765
-17.8088916 -17.80851196 -17.8091336 -17.80969381 -17.81005218
-17.81024653 -17.81076479 -17.81065118 -17.81052082 -17.81053835]
Cluster 2 centroid: [-16.03015213 -16.07135623 -16.08704662 -16.25262547 -16.09607941
-16.09642784 -16.29094356 -16.29314477 -16.30028374 -16.30047067
-16.3007379 -16.3011064 -16.30086728 -16.30035423 -16.28019496
-16.10735534 -16.09986884 -16.11254071 -16.27567611 -16.25449119
-16.08781664 -16.08183359 -16.0584961 -16.16772786 -16.11841847
-16.13296052 -16.14981709 -16.12407941 -16.12291992 -15.97712915
-16.00325719 -16.17656703 -16.03299219 -15.97081329 -15.99680348
-16.00654071 -15.99863707 -16.00306134 -16.11561767 -15.97926701
-16.01109213 -16.19510997 -16.21574883 -16.23145126 -16.21843319
-16.18321392 -16.21998439 -16.27734865 -16.27553246 -16.30051974]
Cluster 3 centroid: [-21.7598124 -21.77808592 -21.79160413 -21.79004622 -21.78741707
-21.7854758 -21.79773246 -21.80917074 -21.80873301 -21.81333605
-21.81486134 -21.81069059 -21.81085916 -21.81081566 -21.81263187
-21.81232735 -21.81609842 -21.81502175 -21.81559815 -21.80641109
-21.79040239 -21.77937738 -21.78125884 -21.78795813 -21.79478793
-21.79875204 -21.80725394 -21.81273518 -21.81374932 -21.81766721
-21.81367863 -21.81324361 -21.81733823 -21.81647091 -21.81664763
-21.81485046 -21.81623165 -21.81685971 -21.81539424 -21.81693583
-21.8144807 -21.81432844 -21.81641109 -21.81498369 -21.81709081
-21.81611202 -21.81544861 -21.8153888 -21.81507341 -21.81716422]

```

图 2-2 光纤电信号四个簇的中心点

经过对”opticalpower”的 K-means 聚类分析，我们无监督的分出了四个簇，中心点分别徘徊在-18、-17、-16、-21 左右。根据以上结果方便我们进一步对其进行数据挖掘。

2.3 频谱分析

刘影^[3]在生命特征提取里使用了离散小波变换，受此为启发，我们决定先对光纤维电信号进行傅里叶变换和小波变换。因为每个 json 的电子信号序列有其相对应的傅里叶频谱图，我们选择图 2-3 所示逻辑进行处理。

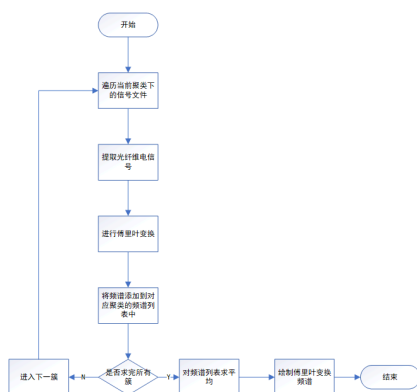


图 2-3 “平均”傅里叶频谱流程图

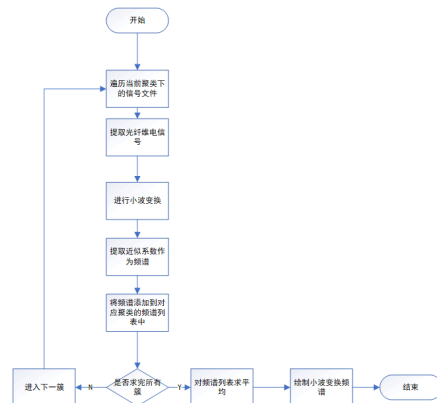


图 2-4 “平均”小波频谱流程图

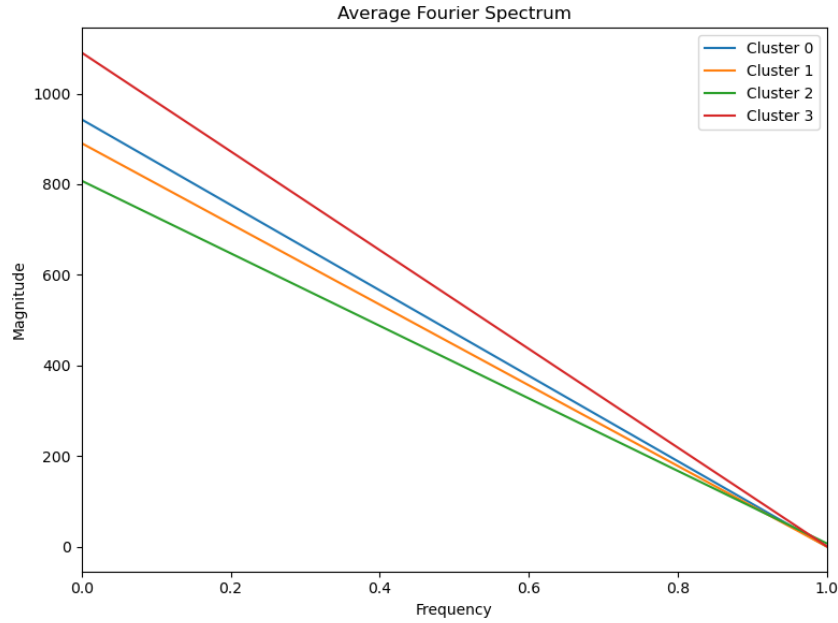


图 2-5 四个簇的“平均”傅里叶频谱图

通过变换得到四个簇的“平均”傅里叶频谱图。其中横坐标(频率/Frequency)表示信号中各个周期性成分的重复次数。在傅里叶变换中,频率范围从 0 到采样频率的一半,对应信号的低频到高频成分。通常以归一化的频率表示,取值范围在 0 到 1 之间。纵坐标(幅度/Magnitude)表示信号在对应频率下的能量或振幅大小。它表示了信号在不同频率上的强度或贡献程度。较高的幅度值表示对应频率上的频率成分在信号中更为显著或重要。

通过绘制傅里叶频谱图,我们可以观察到不同聚类下的平均傅里叶频谱,从而了解每个聚类中信号的频域特征。横轴的频率范围和纵轴的幅度大小会根据数据的采样率和所选的信号长度而有所不同。由图 2-5 知低频信号的特征较明显,而小波变换使用不同的小波基函数来表示信号,如 Daubechies 小波和 Symlet 小波,对于捕捉低频信息表现良好^[4],所以使用 Daubechies 小波变换进行处理,流程和“平均”傅里叶频谱类似。

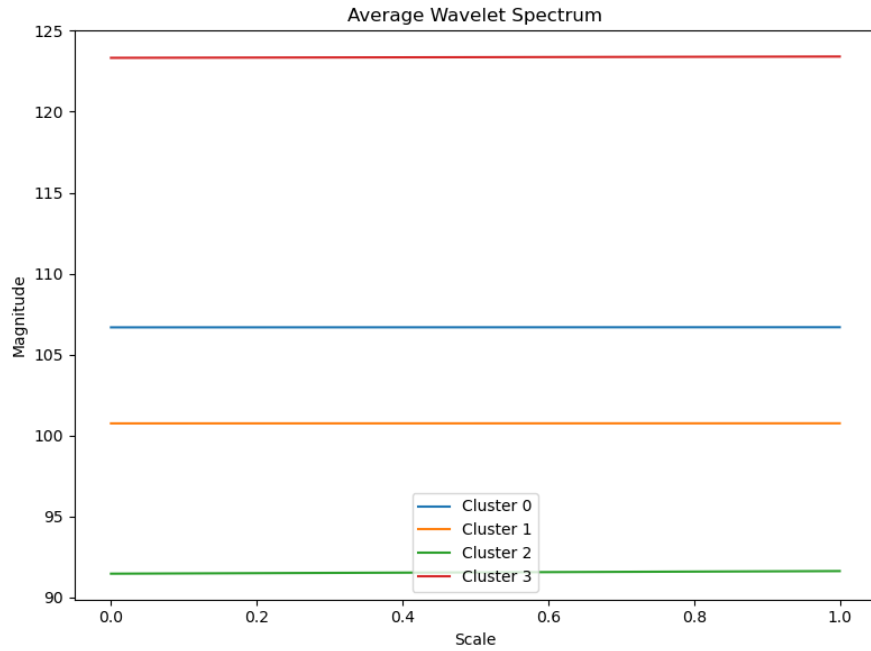


图 2-6 四个簇的“平均”小波频谱图

通过绘制小波变换图，我们可以观察到不同聚类下的平均小波频谱，从而了解每个聚类中信号的频域特征。图 2-6 中横坐标(尺度/Scale)表示小波变换中用来表示频率的参数，它与信号的频率成反比关系，较小的尺度对应较高的频率，较大的尺度对应较低频率。尺度值越大，表示观察到的信号的更低频成分。

纵坐标（幅度/Magnitude）表示信号在对应尺度下的能量或振幅大小,体现了信号在不同频率上的强度或贡献程度，较高的幅度值表示对应尺度上的频率成分在信号中更为显著或重要。综上我们得出，小波频谱图的结果表明每簇的数据有趋同性，不同簇之间的信号有明显差别。

2.4 皮尔逊全相关性分析

经过小波变换后的不同簇的信号特征十分明显。我们先完成前面未完成的相关性分析，思路借助“平均”思想，对每簇数据的光纤维电子信号进行求平均然后求相关性，以及对小波频谱的两个系数和傅里叶变换进行处理后求平均然后求相关性。分析得到相关性矩阵热力图如图 2-7 所示，其中“bt”、“h_t”、“t_M”、“a_o”、“a_o_wl_CA”、“a_o_wl_CD”、“a_o_fft”分别代表“breath”、“heart_rate”、“totalMotion”、“average_opticalpower”、“average_opticalpower_waveletCA”、“average_opticalpower_waveletCD”、“average_opticalpower_fft”。

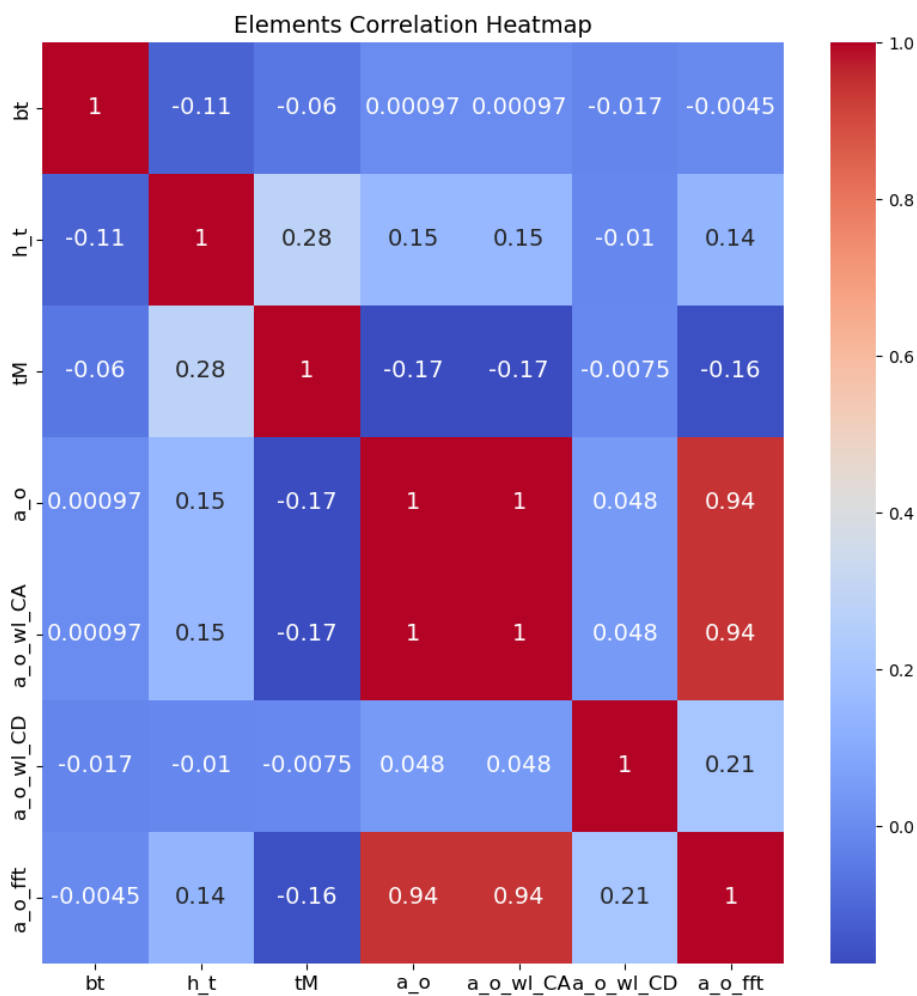


图 2-7 七元素相关系数矩阵热力图

除图 2-1 已分析出的相关性以外，我们可以看到，” heart_rate” 和 ” average_opticalpower” 的相关性最高；” breath” 和 ” average_opticalpower” 的相关性最高；” totalMotion” 和 ” average_opticalpower_waveletCD” 的相关性最高。

2.5 簇分析

针对聚类后呼吸、心率、体动三种生命体征进行基础的数值分析，下示三图分别为各簇的平均值、方差和分布。

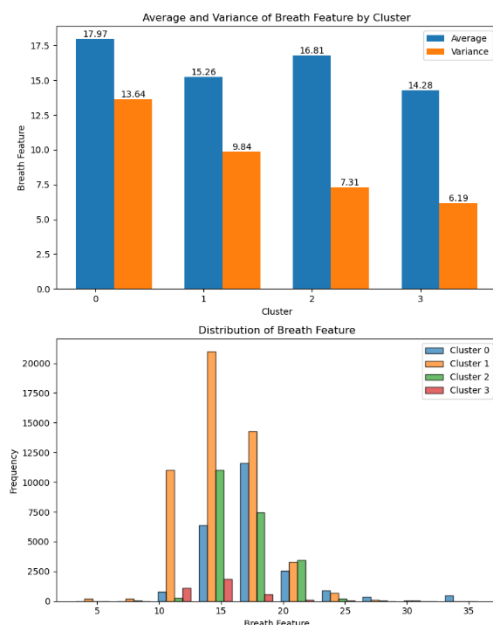


图 2-8 呼吸特征图

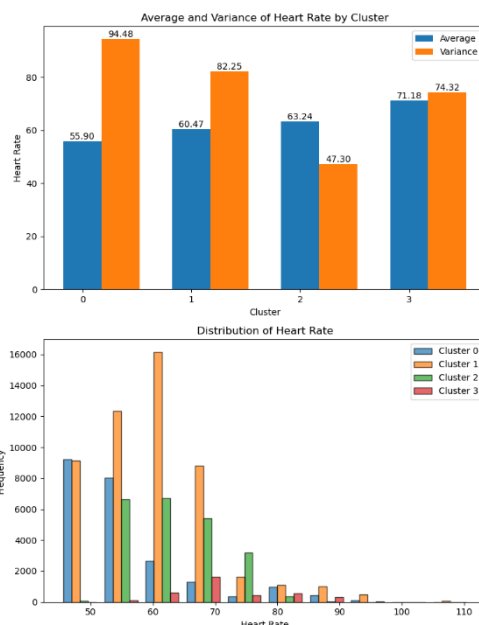


图 2-9 心率特征图

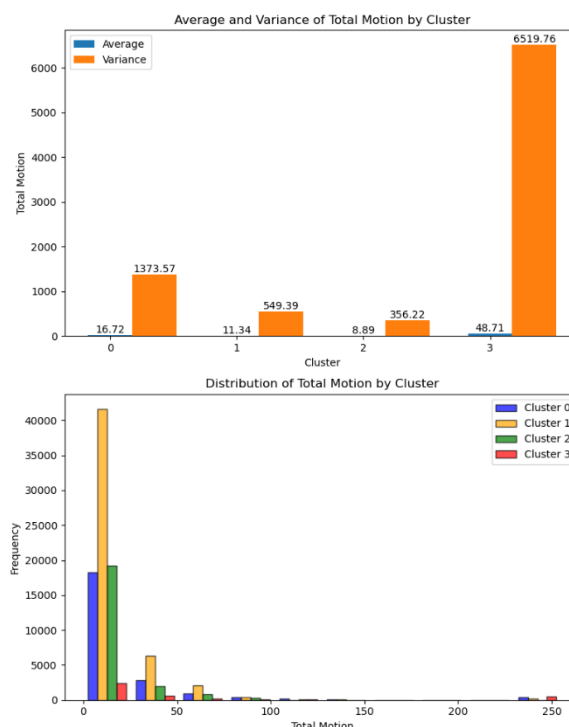


图 2-10 体动特征图

我们观察到呼吸和心率样本在不同簇内的分布有相似性，故使用拟合概率密度函数(PDF)对每个聚类的呼吸和心率的分布进行拟合，如图 2-11 所示。拟合结果表明，在已知数据的样本集中，cluster_1 和 cluster_2 的呼吸和心率的值在对光

纤维电信号使用 K-means 进行聚类后，分布模式相似。由此我们推测光纤电信号对呼吸和心率的模型也是相似的。

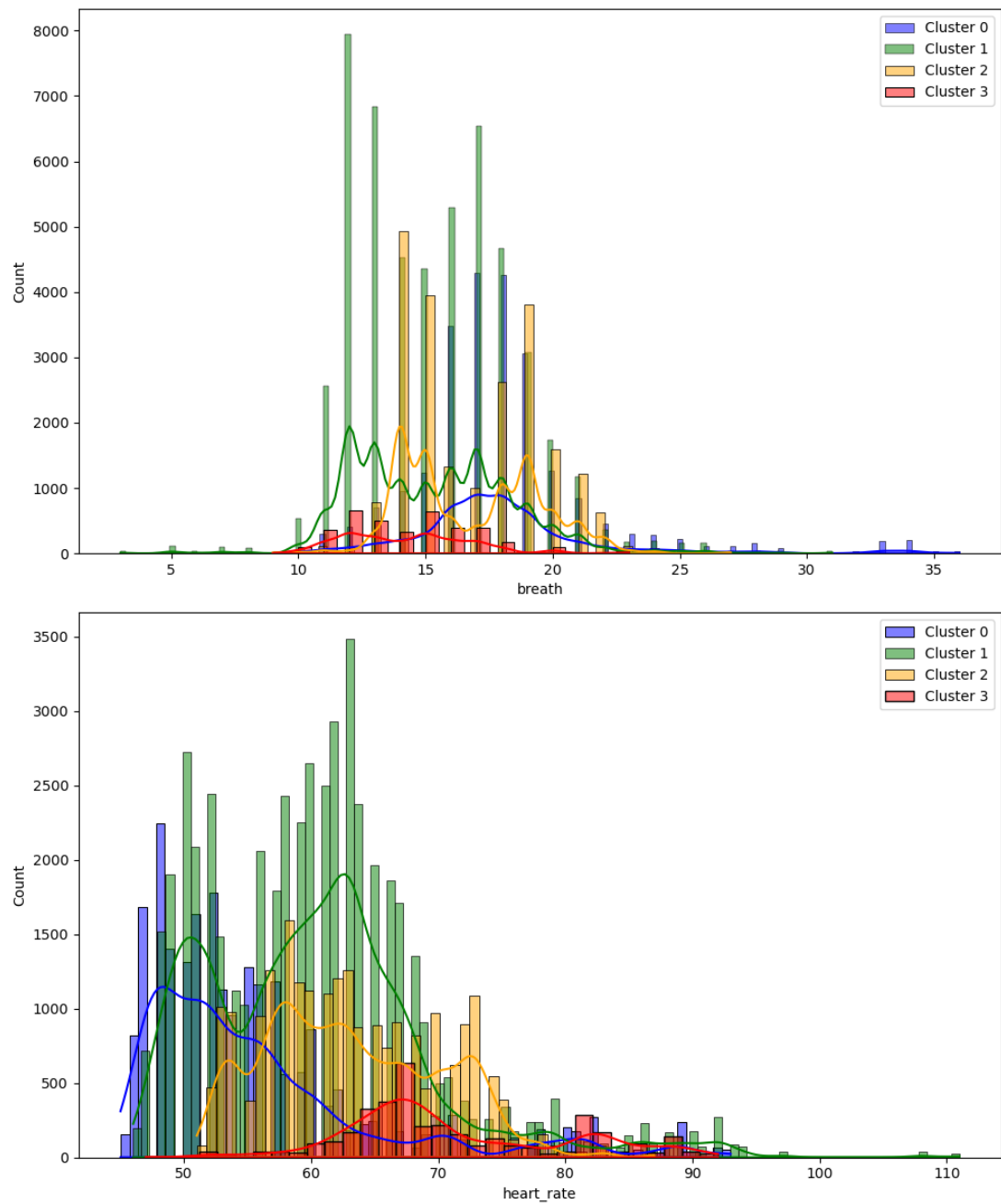


图 2-11 呼吸和心率分布

第三章 模型建立

3.1 电子信号估计生命体征模型建立

经过上一部分的数据处理，我们决定先对” opticalpower” 和” breath” 进行建模。因为输入的是离散的光纤电信号，而输出的是一个确定的呼吸的值，所以这在机器学习中属于“回归”任务。

回归任务是一种监督式机器学习问题，其目标是建立一个函数或模型，将输入特征 x 映射到连续型目标变量 y 的预测值。我们可以将回归任务表示为一个函数关系 $y = f(x) + \varepsilon$ ，其中： x 是输入特征向量，包含多个特征 x_1, x_2, \dots, x_n ，表示待预测的样本的属性； y 是连续型目标变量，表示我们希望预测的数值； f 是未知的真实函数，表示特征与目标变量之间的关系； ε 是表示模型预测的误差或噪声，即模型无法完全准确地预测真实目标变量的部分。

回归任务的目标是通过观察和学习一组已知的训练样本 (x_i, y_i) ，其中 $i = 1, 2, \dots, m$ ，来估计或逼近真实函数 f 。我们的目标是找到一个函数或模型 $h(x)$ ，使得它能够最好地拟合已知训练样本，并能够在未见过的测试样本上进行准确的预测。

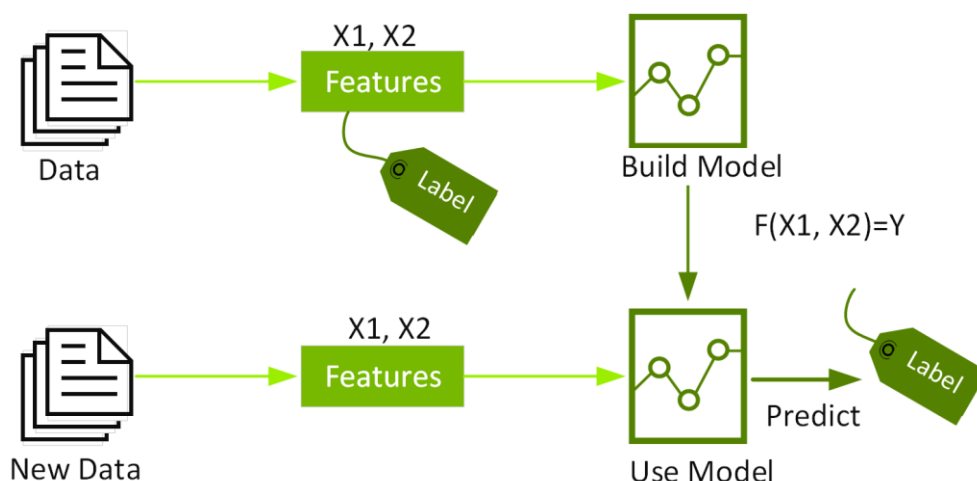


图 3-1 监督式机器学习^[5]

而处理回归任务通常有以下几种算法：线性回归，支持向量机（SVM），随机森林，逻辑回归。但是逻辑回归一般用于二分类问题，这里并不适用，所以这里我们对前三种模型都进行了建立。其中线性回归是将每个文件的光纤电信号进行求平均来用来拟合。

线性回归是一种常见的回归算法，用于建立输入特征 x 与连续型目标变量 y 之间的线性关系模型。它假设特征与目标之间存在线性关系，并试图找到最佳拟合直线来预测目标变量。

线性回归模型的数学表示为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

其中： y 是目标变量（也称为响应变量或因变量）； x_1, x_2, \dots, x_n 是输入特征（也称为解释变量或自变量）； $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ 是回归系数（也称为权重或斜率），

表示特征对目标的影响程度； ε 是误差项，表示模型无法完全准确预测目标变量的部分。

线性回归的目标是找到最佳的回归系数，使得模型预测值与实际观测值之间的误差最小化。通常使用最小二乘法来估计回归系数，该方法通过最小化观测值与模型预测值之间的残差平方和来找到最佳拟合直线。

随机森林算法^[6]基本原理是采用 Bootstrap 子自采样的方法获得不同的样本集用于构建模型，从而增加了模型之间的差异度并提高了外推预测的能力。对于属性选择：首先，从基决策树结点的属性集合中随机选择一个包含 k 属性的子集（ k 通常为 $\log 2d$ ，其中 d 为属性集合的大小）；其次，从这个子集中选择一个最佳属性用于划分；最后，采用多数结果法、平均法、投票法等综合考虑多个决策树的预测结果确定最终随机森林算法的预测结果^[7]。

支持向量机^[8]由 Cortes 与 Vapnik 在 1995 年第一次被提出，在近些年得到了很大发展，是基于结构风险最小化原则的一种统计学算法。SVM 对小样本集有着优异的学习性能，由刚开始处理分类问题成功地扩展到了解决回归的问题，并不再依据传统方法中应用的经验风险最小化原则，而是依据结构风险最小化原则，使其拥有了更好的泛化能力。SVM 需要设置的参数不多，平衡了模型的知识性与复杂性，而且训练问题整体可以归纳为一个求解凸二次规划的问题，确保了所求得到的解是全局最优解，往往也是唯一的解^[9]。

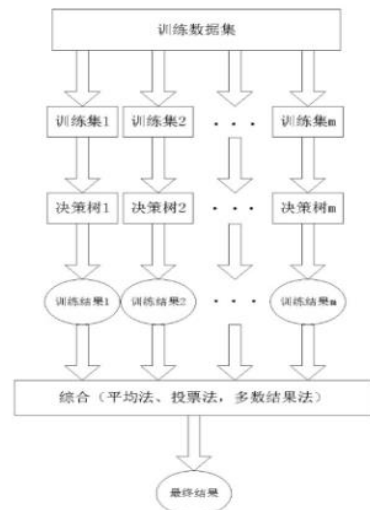


图 3-2 随机森林算法原理

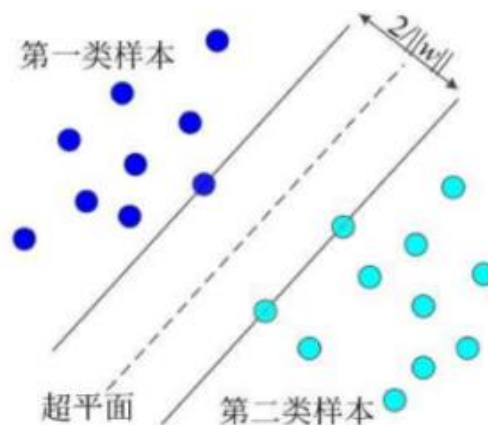


图 3-3 支持向量机线性二分类^[10]

除此以外，我们还使用了 XGBoost^[11]。XGBoost（eXtreme Gradient Boosting）是一种基于梯度提升树的机器学习算法，它在许多机器学习竞赛和实际应用中表现出色。XGBoost 结合了梯度提升框架和正则化技术，通过迭代训练一系列弱学习器（通常是决策树），以最小化损失函数并提高预测性能。

梯度提升树是一种集成学习算法，通过迭代训练多个弱学习器来构建一个强大的预测模型。梯度提升树的核心思想是通过逐步优化损失函数的梯度来改进模型的预测能力。

每个弱学习器都尝试纠正前一个学习器的错误，并在残差之间构建关联，以逐步减少模型的预测误差。弱学习器是指预测性能相对较弱的学习算法或模型，

如决策树。XGBoost 使用决策树作为基本的弱学习器，通过组合多个决策树来构建更强大的模型。

梯度提升是指通过优化损失函数的梯度来更新模型的参数或权重，以逐步改进模型的预测能力。在每个迭代步骤中，XGBoost 根据残差计算梯度，并使用梯度信息来训练下一个弱学习器。

正则化是为了避免过拟合，即模型过于复杂而在训练数据上表现良好但在测试数据上泛化能力较差。XGBoost 使用正则化技术（如 L1 正则化和 L2 正则化）对模型进行约束，防止过度拟合并提高模型的泛化能力。

3.2 针对不宁腿综合征的电子信号序列分析

对于第三问，我们选择用不宁腿综合征^[12]为例进行分析。

此类患者常常因为失眠去医院就诊，但多有特征性临床症状，主要表现为夜间睡眠时双下肢不适感，迫使不停活动下肢、捶打小腿或下地行走，进而影响患者的睡眠。虽然对生命没有危害，但很多患者症状持续多年。因为体动值越大，反应了体动幅度的越大，所以以此为切入点进行分析。

经过 2.5 章节的簇分析，再根据体动特征图，我们发现 cluster_3 的体动值最大，且方差最大，表明在 cluster_3 中，老年人“活动”的幅度较大，且次数较别的簇也更多。所以我们针对 cluster_3 的电子信号序列进行分析。首先进行的是电子信号平均数和中位数的分析。

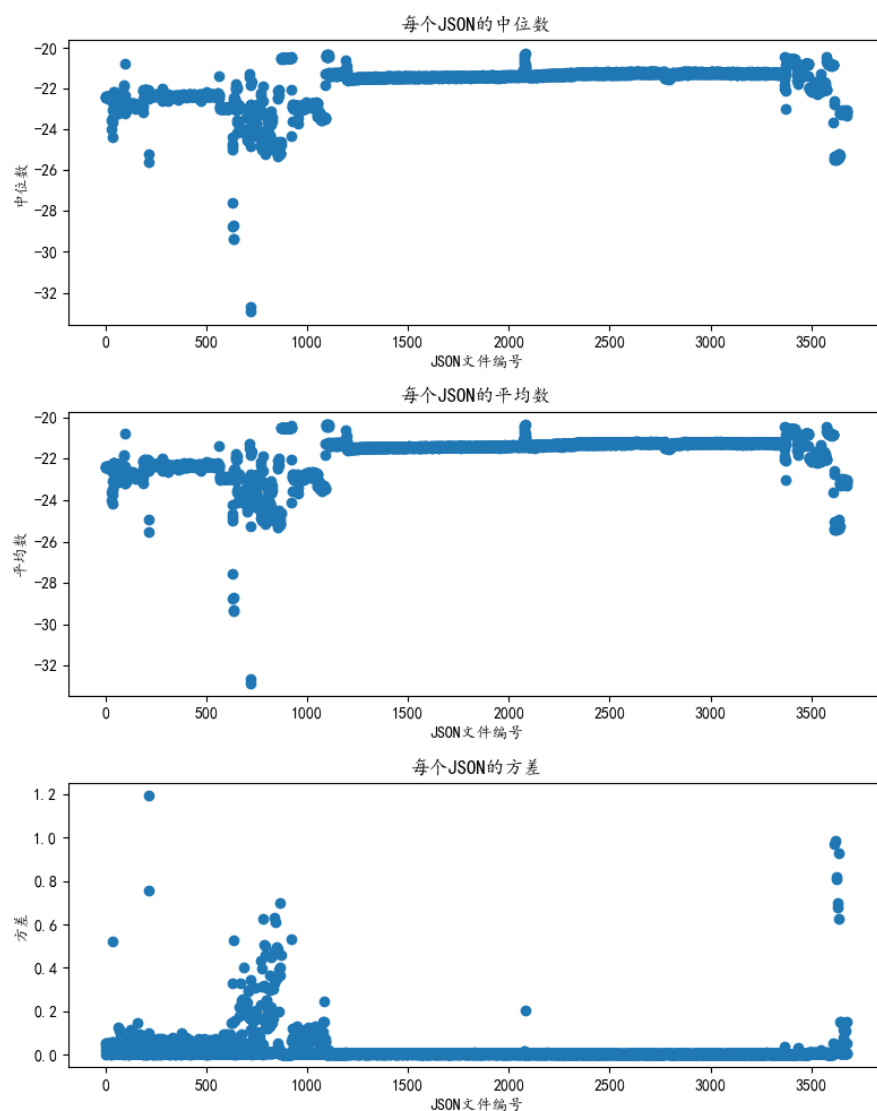


图 3-4 电信号数据分析

分析后发现绝大多数电信号的平均数和中位数都在-22 上下，所以推测在-22 左右时身体处在静息状态下。而方差大于等于 0.5 时，身体则处于非静息状态，即有“活动”的。

针对方差在 0.5 以上的电子信号序列，我们将其提取出来，到新的目录进行分析。分析后发现这些电子信号序列的平均值多在-24 以下，与-22 差值较大，故可认作是有“活动”的。我们发现所有高方差数据均是第 28 号目标的数据，故下面对他进行分析。

表 3.1 高方差序列的数据

dataname	opticalpower	totalMotion
28_2023-05-24-06-04-41.json	-24.1556	2
28_2023-05-24-06-10-04.json	-24.9122	0

28_2023-05-24-06-10-06.json	-25.545	0
28_2023-05-25-05-25-55.json	-22.6038	0
28_2023-05-25-05-29-35.json	-24.7502	48
28_2023-05-25-05-29-51.json	-24.7086	0
28_2023-05-25-05-31-04.json	-24.757	48
28_2023-05-25-05-31-13.json	-24.7112	48
28_2023-05-25-05-31-45.json	-24.9202	19
28_2023-05-25-06-10-08.json	-24.099	33

用第二章中求分布的方法对第 28 号目标的体动值数据进行独立分析，分析得图 3-5。在 28 号目标的体动值中，33 和 48 都属于出现频率较低的体动值，我们可以认为当 28 号目标的体动值超过 33 时，28 号表现出有“活动”的趋势，可以认为是不宁腿综合征的发病。但同时要知道的是，如果目标在一分钟内的电子信号序列的方差偏低，则可能是目标在这一分钟内是活动的“过程”，因为“活动”可以持续一段时间，而电信号会因此发生改变，“活动”持续的时间越久，电信号序列波动的幅度就越小，导致方差偏低。

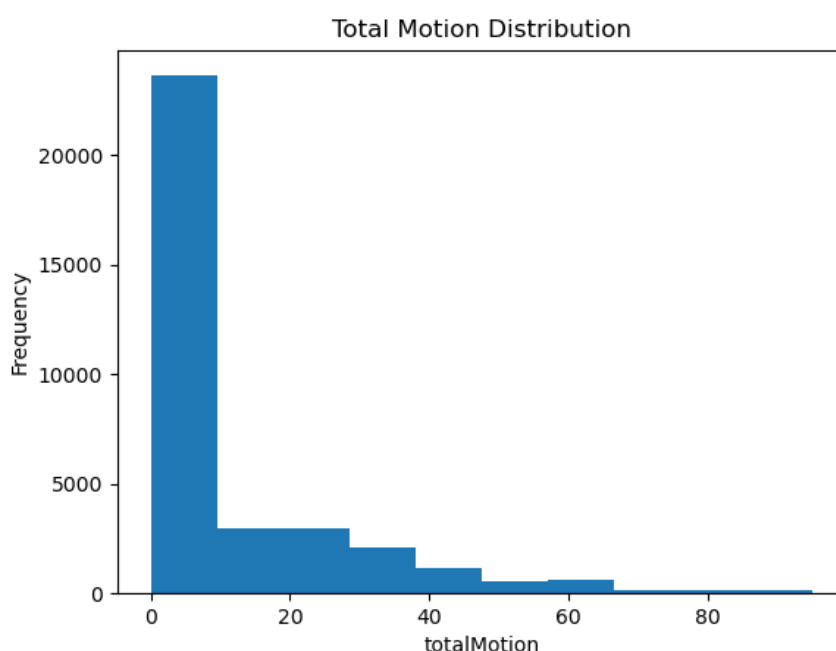


图 3-5 28 号目标的体动值分布

3.3 模型评估

建立了四种模型，所有模型的训练集和测试集的比例为 8：2。SVM 模型使用来自 sklearn.svm 的 SVR 模型；Random Forest 使用 sklearn.ensemble 的 RandomForestRegressor 并构建决策树数量为 100，也就是随机森林中的决策树个数，为了使结果可复现，随机性的种子值也已固定；Linear Regression 使用 sklearn.linear_model 的 LinearRegression 包；XGBoost 使用 xgboost 库的 XGBRegressor。训练结束后，进行模型的性能评估，我们建立的是根据光纤电信号输出生命体征的模型，所以每个模型输出的是确切的值。

因为最终输出的是一个值，选择的性能评判指标是均方误差(MSE)，R 系数和均方根误差(RMSE)。

表 3.2 光纤电信号预测呼吸不同模型性能比较

Model	R	RMSE	MSE
SVM	0.2742	2.8742	8.2612
Random Forest	0.4212	2.5669	6.5887
Linear Regression	0	3.3741	11.3844
XGBoost	0.3941	2.6262	6.8968

根据 2.5 的结论，我们对心率和体动进行了相似的建模。其中心率模型的性能表现和呼吸模型的性能表现都比较优秀，也印证了 2.5 的结论，电信号-呼吸和电信号-心率的模型是类似的。

表 3.3 光纤电信号预测心率不同模型性能比较

Model	R	RMSE	MSE
SVM	0.2536	8.1592	66.73622
Random Forest	0.5786	6.1380	37.6750
Linear Regression	0.0260	9.3326	87.0981
XGBoost	0.5555	6.3044	39.7460

表 3.4 光纤电信号预测体动不同模型性能比较

Model	R	RMSE	MSE
SVM	-0.1079	33.5659	1126.6676
Random Forest	0.2373	27.8500	775.6231
Linear Regression	0.0335	31.3497	982.8046
XGBoost	0.2021	28.4854	811.4185

结果表明，根据 optical_power 可以很好的预测出呼吸，一定程度上可以预测心率，而对于体动值无法很好的预测。

3.4 模型改进

根据 2.1 得出的结论，体动值和心率有比较大的相关性，以此为切入点，我们决定尝试将光纤电信号——心率模型和心率——体动模型进行串联。逻辑上类似神经网络的神经元。



图 3-6 信号神经网络神经元

因随机森林在该数据集上表现优异。心率——体动模型选择随机森林和线性回归进行性能比较。

表 3.5 心率预测体动不同模型性能比较

Model	R	RMSE	MSE
Random Forest	0.2181	28.1978	795.1147
Linear Regression	0.0827	30.5413	932.7713

表 3.6 组合模型性能

Model	R	RMSE	MSE
Combined	-0.0842	32.4521	1053.1406

根据表 3.5 可以看出，随机森林在心率——体动模型上表现较好，故拿其与光纤电信号——心率模型中性能最优异的模型进行组合。再进行性能评估。虽然组合模型的性能不如单独使用随机森林进行预测的效果好，但是依旧比 SVM 和线性回归的性能好。SVM 模型的训练时间最久，花费成本最大；线性回归训练时间最短，性能也最差；组合模型的训练时间只比单独训练随机森林的时间长，性能却比 SVM 和线性回归都好。在这种情况下，可以对不同的机器学习模型进行组合尝试寻求最优解，因为篇幅问题，本论文不在此进行求最优解的过程。

第四章 需要被关注的生命体征

近年来,随着社会经济的发展和人们生活水平的提高,人们对自身身体健康的关注度也越来越高。

4.1 体温监测

体温是最基本的生理参数之一,对于疾病的发现和预防有重要的作用,从过去疫情防控中体温监测执行情况来看,体温监测工作需要专业人员实时进行测量登记,消耗大量人力且存在时间间隔,做不到实时监控,监测空挡内的体温异常往往来不及发现并及时处理,所以体温的实时监控是有必要的。若有无线传输体温监测技术的介入或可以有效的改变目前实践中的体温监测方式^[13],从而实现高效的实时掌控。监测系统主要负责的模块有体温采集、数据处理、实时呼叫和电量预警。

监测过程可使用电子体温传感器,该传感器的优势为传感器固定在被测对象体表,不影响受试者正常活动,相比于传统的水银温度计可以避免消毒不彻底造成的交叉感染和测温时间长的问题^[14],红外体温计虽能快速响应且能避开交叉感染但是测温者可能操作不当致使读书不精确从而影响判断^[15, 16]。

新进^[17]提出并完成医疗器械注册检验的基于 Lo Ra 无线传输技术搭建的无线局域网,扩展了无线体温监测系统的覆盖面,使得医院整个院区多栋住院大楼成为一个监测整体,方便服务器布置和数据管理,有效减少了医护人员的工作量和工作时间,大大提高了医护人员的工作效率。

4.2 血压监测

高血压疾病是临床上常见的疾病,这种疾病发病率较高。患者发病后临床上如果不采用积极有效的方法治疗将会诱发心血管疾病,如:心律失常、冠心病、心力衰竭等,影响患者治疗预后和患者生活质量^[18]。根据世界卫生组织发布的《2018 世界卫生统计报告》显示,在 2018 年大约有 4100 万人死于非传染性疾病(NCDs),死于心血管疾病的人数为 17900 万,占据总 NCD 人数的 44%,其中大约 940 万人死于高血压引起的并发症^[19]。

大于 70 岁者,其一生中患高血压的几率高达 90%以上^[20]。因此,大多数人一生中或迟或早都需要进行 1 次或多次 24h 动态血压监测。但因服务资源有限,且 24h 动态血压监测费用较高、耗时耗材以及被监测者由于短间隔的反复佩戴器械测量血压会产生不适,只能选择极少数人,在最需要的时候,进行监测。2011 年英国高血压指南明确建议,诊室血压在 140~179/90~109mm Hg(1mm Hg=0.133kPa)范围内的 1、2 级高血压均需经 24h 动态血压监测确诊后,才能启动降压治疗^[21]。所以动态血压监测可以准确反映一个人的血压风险,显著提高高血压的诊断能力,提高心血管风险评估的准确性。

对于动态血压监测(ABPM),从我国该研究方向的发文数量来看,虽然起步较晚,发文总量较少,但我国已有学者成为核心作者^[22]。

参考文献

- [1] 日内瓦:世界卫生组织. 世卫组织关于身体活动和久坐行为的指南 [WHO guidelines on physical activity and sedentary behaviour]. [M/OL]. 2021[(Chinese)].
- [2] SHEN C-P, KAO W-C, YANG Y-Y, et al. Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines [J]. Expert Systems with Applications, 2012, 39(9): 7845-52.
- [3] 刘影, 董升华. 基于可视图法的多载波生命体征监测算法 [J]. 云南大学学报(自然科学版), 2022, 44(06): 1135-45.
- [4] A. BOGGESS F J N. A first course in Wavelets with Fourier Analysis [M]. 2 ed., 2001.
- [5] NVIDIA. XGBoost [Z]. <https://www.nvidia.com/en-us/glossary/data-science/xgboost/>
- [6] 董红瑶, 王弈丹, 李丽红. 随机森林优化算法综述 [J]. 信息与电脑(理论版), 2021, 33(17): 34-7.
- [7] TIN KAM H. Random decision forests; proceedings of the Proceedings of 3rd International Conference on Document Analysis and Recognition, F 14-16 Aug. 1995, 1995 [C].
- [8] 时安琪. 基于 SVM 的矿井提升机载荷辨识研究 [D]; 太原理工大学, 2017.
- [9] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-97.
- [10] 卓信概. 基于新型支持向量机的轴孔装配接触状态识别方法 [D]; 浙江工业大学, 2020.
- [11] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System [Z]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA; Association for Computing Machinery. 2016: 785 - 94. 10.1145/2939672.2939785
- [12] 马建芳, 陈生弟. 不宁腿综合征的诊断和治疗 [J]. 中国神经免疫学和神经病学杂志, 2010, 17(5): 379-81.
- [13] 杜江波. 无线传输体温监测技术探究 [J]. 数字技术与应用, 2020, 38(08): 174-5.
- [14] 胡红波, 袁田, 谢涛, et al. 防控新型冠状病毒肺炎体温连续监测预警系统的设计 [J]. 中国医学装备, 2021, 18(01): 137-40.
- [15] 顾俊俊, 陈向东, 丁星, et al. 一种快速响应的红外实时体温监测系统 [J]. 物联网技术, 2023, 13(01): 19-23.
- [16] 汪洪军, 柏成玉, 王景辉, et al. 红外测温仪在应对新冠疫情中的应用 [J]. 中国计量, 2020, (05): 17-20.

-
- [17] 黄鑫. 基于 LoRa 无线传输技术的体温监测系统的设计 [J]. 数字技术与应用, 2022, 40(12): 183-5.
- [18] 陈海英. 老年高血压引发心血管疾病的相关因素分析及临床对策 [J]. 中国继续医学教育, 2015, 7(06): 30-1.
- [19] 吴金桐, 桑海峰, 徐州, et al. 实时血压监测系统设计 [J]. 电子测量技术, 2019, 42(24): 63-70.
- [20] 张颖, 杨冬梅, 王佳. 短时间动态血压监测对健康人群体检的意义 [J]. 中国煤炭工业医学杂志, 2013, 16(07): 1174-6.
- [21] 王继光. 24h 动态血压监测:适用人群及临床意义 [J]. 中华高血压杂志, 2014, 22(07): 614-5.
- [22] 丁雯, 张雪芳, 陈雯. 基于 Citespace 的动态血压监测研究热点可视化分析 [J]. 心脏杂志, 2022, 34(01): 49-54.