

第六届湖南省研究生数学建模竞赛

题目 基于层次分析法和机器学习的黄桃品质 评估和贮藏分析

摘 要：

黄桃商用价值较高，但是容易腐烂，需要精准把握其品质评估方法和存储方式才能更好地采摘、存运与售卖，推动黄桃产业的数字化转型。因此，准确评估黄桃品质和制定黄桃存储策略具有重大意义。本文通过题目给定的数据集，对黄桃品质分析以及变化情况建立了数学模型，完成了对人工评判和机械指标的一致性评判与人工评分等级的合理预测，给出了黄桃鉴别和贮藏策略。

针对问题一，我们首先采用统计方法和加权主成分分析对两类指标关系进行了可视化分析，然后以每批桃为最小研究单位建立了人工评价体系和机械评价体系的映射，从而可以在同一空间上比较两个体系的一致性。采用的映射方法为线性回归和神经网络，其中神经网络建立了更加符合实际分布的映射从而取得了更好的效果。在建立了映射后，我们选取两个体系评优的交集作为评优的范围从而得到了较为严格的品质分析定量模型。然而上述体系只用到了部分指标，而其余指标（如芳香、桃子颜色）的确是黄桃品质的重要判断依据。为了利用尽可能多的指标，以及衡量不同指标的重要性程度，我们采用层次分析法进行模型构建。得到模型符合人工评分要求在总体上高于机械指标的认知，我们的模型对于优质的认证也比专家经验更加严格。

针对问题二，我们提出了一种多模型联合预测机制来应对数据量较少出现的模型过拟合或欠拟合问题，分别通过线性回归模型、支持向量机（SVM）分类模型、神经网络回归模型对数据进行拟合预测，再通过加权投票机制得到最终的预测结果。在回归模型中，我们利用样本的机械指标预测出人工评分，再利用人工评分标准进行优劣分类，在分类模型中，我们利用 SVM 对机械指标进行二分类。

针对问题三，我们以每一批黄桃为研究对象得到了黄桃在不同环境品质变化情况，采用基于原理的微分方程和基于数据的函数拟合对黄桃品质在不同环境随存储时长变化情况进行了模型建立。一方面，黄桃的品质变化其根本原因是内部化学反应，因此构建化学动力学模型来进行品质变化预测是合理的。另一方面，也可以通过数据直接拟合黄桃指标随时间变化趋势达到预测的目的。首先，我们采用了样条插值的方法对单变量进行拟合，然后采用长短记忆网络进行对多指标的同时拟合和预测，并且得到了预测次数和预测均方误差的关系图。通过对模型的分析，我们的发现如果想保持较好的风味，可以在冷柜中贮藏 8-14 天；如果想延长贮藏时间，可以在冷库中贮藏 20 天左右，这与普遍认知是一致的。

关键词：黄桃品质评估与预测 主成分分析 层次分析法 线性回归 机器学习

1 问题重述与分析

1.1 问题重述

本题的核心任务在于利用机械指标和人工评分，综合评判该品牌黄桃果实的品质，以满足人们对于高品质黄桃的需求。分析题目和附件数据，将得到的信息整理如下：

机械指标：

2018 年机械指标：带皮硬度、去皮硬度、TSS、色差。

2019 年机械指标：带皮硬度（右）、带皮硬度（左）、去皮硬度（右）、去皮硬度（左）、TSS（右）、TSS（左）、单果重、纵径、横径、侧径、色差（皮左）、色差（皮右）、色差（肉左）、色差（肉右）。

2020 年机械指标：带皮硬度、去皮硬度、TSS、色差（皮）、色差（肉），冷库贮藏条件下无带皮硬度和色差指标。

人工指标：

2018 年人工指标：色泽、质地（硬度、汁液丰富等）、香（芳香）、味（糖酸、可固）。

2019 年人工指标：色泽、质地、芳香、味道。

其中 2019 年人工测试员相较 2018 年的人工测试员更为专业，评分更严格。

贮藏环境：

冷柜（8℃）、冷库（1℃）。

专家经验的评测标准：

机械指标评测“优质”的度量标准：果实硬度 $> 1.5\text{kg}/\text{cm}^2$ ，TSS 含量下降率 $< 20\%$ 。

人工指标评测“优质”的度量标准：色泽、质地、味道三个指标均分 ≥ 80 分。

普遍认知：

1) 机械指标易满足，人工评分要求更高。

2) 冷库（1℃）贮藏条件下，黄桃风味下降，保质期较长。冷柜（8℃）贮藏条件下，黄桃风味保持良好，但保质期较短。

评测样本取样方式:

表 1 样本取样方式

贮藏环境	年份	取样 间隔 时间 /天	每批抽取次数/次	每次抽取样本数/个	备注
冷柜 (8℃)	2018	3	5 (6.11、6.14、6.16)	16 (6.11、6.14、6.16)	括号中日期代表在该天进行贮藏的批次, 如: (6.14) 代表在 6 月 14 日进行贮藏的那一批次。
	2019	2	9 (6.14)、 11 (6.17)、 9 (6.20)。	初始第一次: 8 (6.14)、 10 (6.17)、 10 (6.20), 之后每次 5 个	
	2020	2	9 (6.14)、 10 (6.18)、 9 (6.22)。	初始第一次: 16 (6.14)、 20 (6.18)、 20 (6.22), 之后每次 10 个	
冷库 (1℃)	2018	5	5 (6.11、6.14、6.16)	16 (6.11、6.14、6.16)	
	2019	2	10 (6.14)、 12 (6.17)、 10 (6.20)。	初始第一次: 8 (6.14)、 10 (6.17)、 10 (6.20), 之后每次 5 个	
	2020	2	6 (6.14、6.18、6.22)	10 (6.14)、 16 (6.18)、 13 (6.22)。	

1.2 问题分析

针对问题一, 问题一有三个要求, 分别是比较指标评分之间的关系、机械指标与人工指标评判一致性和构建黄桃品质分析的定量模型并用该模型评判“专家经验”和“普遍认知”。重点是找到机械指标与人工指标内在的关系, 综合考虑机械和人工的各项指标, 建立科学的定量模型, 判断是否与专家经验和普遍认知一致。

针对问题二, 在 2020 年度的评测数据中只提供了机械指标数据 (包括带皮硬度、去皮硬度、TSS 含量和色差), 没有提供人工评分。问题二的最终目的是利用 2020 年给出的机械指标数据, 判断题中给出的 9 个样本的人工评分等级。重点是通过机械指标与人工指标的关系预测样本的人工评分, 需要考虑数据量少, 数据不对应所带来的问题。

针对问题三, 依据前两问得到的信息和结果, 建立黄桃品质随贮藏环境和贮藏时间变化的预测模型, 需要综合考虑黄桃各项评价指标受采摘时间, 贮藏方式, 贮藏时间等因素的影响。

2 模型假设和数据处理

2.1 问题假设

1. 所有机械指标和人工评分分布满足联合高斯分布。

该假设有利于测试样本的数据生成，并为 Z score 归一化提供支撑。

通过数据统计和观察四分位数以内分布，所有指标的确满足正态分布：

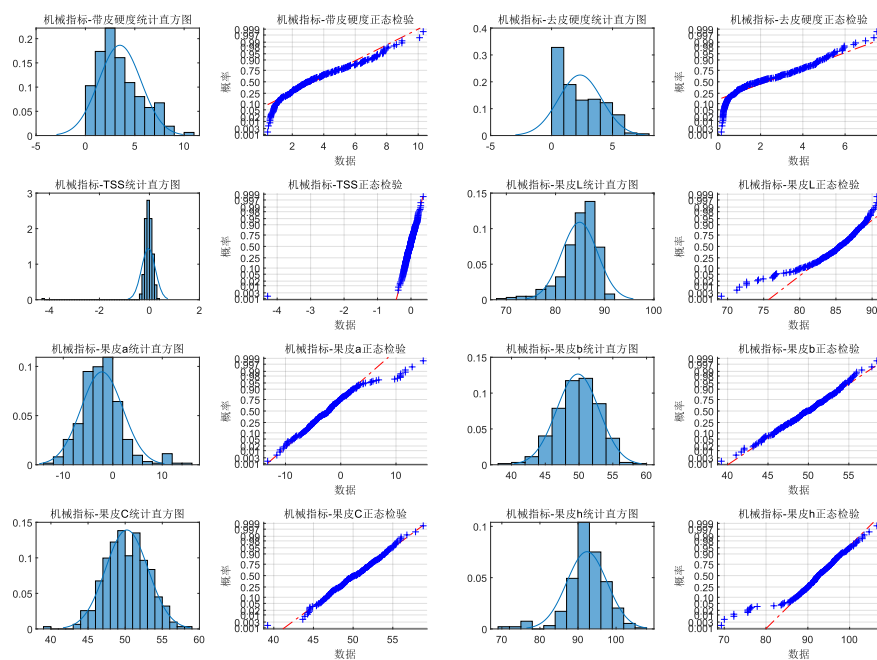


图 1 人工评分指标正态性检验

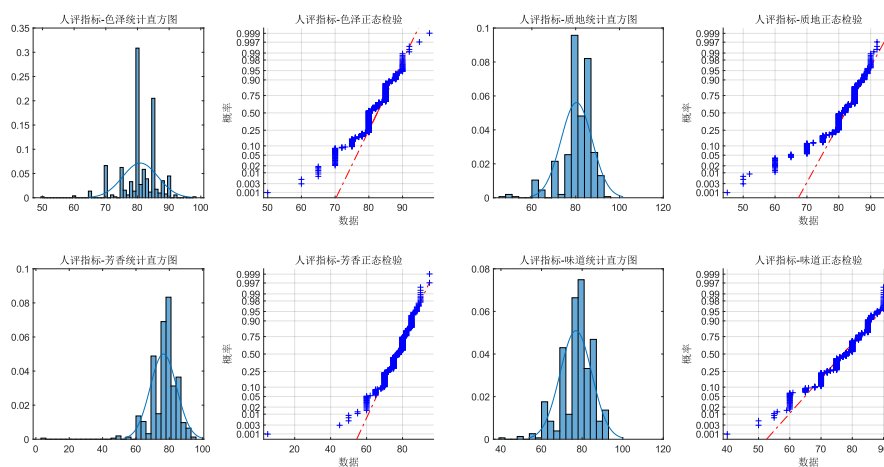


图 2 机械指标正态性检验

2. 不同年份桃子的品质随时间变化的曲线具有相似性

通过该假设就可以把不同年份不同采摘时间相同储存环境下的批次合在一起来探究储存环境对品质变化的特性。

同一地点同一种植基地同一时间段的桃子具有相似性。

3. 同批采摘的桃子测验的结果经过平均后可以代表整批桃子的特性

问题 1 和问题 2 中，通过该假设可以让我们以每一批桃子作为研究对象从而构建每批桃子人工打分和机械指标的联系。同时，该假设也为问题 3 中构建每批桃子品质变化曲线提供可能。平均也可以降低异常值的影响。

根据假设 1，桃子各指标满足联合正态分布，通过取平均可以代表整批桃子属性。

2.2 符号说明

表 2 部分符号说明表

符号	含义
$Score_{man}$	人工指标域
$Score_{mec}$	机械指标域
$F(\cdot)$	线性映射函数
MSE	损失函数
U	色样集
F	综合指标最终得分
Q	品质
$\oplus$	异或
CI	一致性指标
CR	一致性比例
RI	随机一致性指标
$u(\cdot)$	阶跃函数

2.3 数据处理

由于所给黄桃数据涉及采摘时间、贮藏方式、贮藏时间等多个时空信息，而且不同年份的数据组织方式和信息各有不同。如 2018 年的机械指标数据中包含去皮硬度、带皮硬度、TSS、L、a、b、C、h 等指标，2019 年的机械指标相对于 2018 年又增加了桃左右侧的分别测量，以及增加了果实大小（纵经、横经、侧经）等数据，2020 年的数据中又舍弃了果实大小数据。而且如题目中提到，2018 年与 2019 年人工评分的专业与严格程度也不一致。因此为了解决题目中涉及的回归、分类、预测等任务，对数据进行处理是非常有必要的。

我们对数据的处理过程可由下图表示：



图 3 数据处理过程

2.2.1 整理编号说明

为了便于对数据进行筛选处理，我们先对所有数据根据时空进行编号，为不同的黄桃样本添加批所属代码，代码记录其时间、采摘日期、储存温度、测试时间（距离采摘日期的天数）等信息，格式如下图所示：



图 4 数据编号格式

通过对数据进行代码编号，我们可以很方便地对数据进行分类，并且将机械指标与人工指标通过编号对应起来。

2.2.2 异常值处理

通过对数据的统计观察，我们发现在所给黄桃数据中存在一些明显异常的数值，如在“附件 3-2018 年人工打分数据”中 6.16 记录的人工指标黄桃质地评分就存在 875 的异常值；在“附件 5-2019 年黄桃机械指标数据（色差）”中冷库数据 7 月 3 日记录的色差指标 a 值中存在 $-7.2e14$ 的异常值；在“附件 4-2019 年黄桃机械指标数据（TSS）”中 6.20 冷库记录的第六天机械指标 TSS 值中存在 111.5 的异常值等。对于以上的异常值，由于我们无法确定原值多少，而且没批次数据量比较少直接拟合插值容易对模型的拟合能力造成负面影响，因此我们选择直接将这些异常值去除。

2.2.3 TSS 增长率计算

在题目中提到的专家经验评分方法中，利用硬度 ($> 1.5kg/cm^2$) 和 TSS 含量下降率（即 TSS 初始指标和贮藏期间的指标对比，下降率 $< 20\%$ ）两个指标来评价黄桃质量好坏。而所给数据中直接给出的是 TSS 含量，因此我们还需要对 TSS 下降率进行计算。为了方便后期对黄桃质量的机械指标评价，我们将 $TSS_{下降率} < 20\%$ 这一指标改为 $TSS_{增长率} > -20\%$ ，直接计算 TSS 增长率。

显然，对于同一批（采摘时间与贮存方式相同）黄桃来说，不同测试时间用于测试的黄桃样本是不一样的。而要计算 TSS 增长率，我们需要一个统一的初始 TSS 值。我们不妨将每一批黄桃贮存第 0 天测得的 TSS 值取平均作为初始 TSS 值，而后该批每个黄桃样本的 TSS 增长率通过与该初始 TSS 值对比得到。

2.2.4 归一化

在所给数据中，存在数据范围不一致的问题，如在机械指标中，果实硬度一般在 0~10 之间，而 TSS 增长率主要分布 -0.5~0.5 之间。在人工指标中，各项评分又分布在 50~100 之间。量级的不同会给分析工作带来困难，对于后期的拟合、分类、预测任务也会造成不良影响，因此对数据进行标准化处理是十分必要的。

从模型假设中我们合理假设这些指标的分布为正态分布，因此我们对这些自变量都进行 Z-Score 标准化，将其归一化为 (0, 1) 正态分布。

2.2.5 批平均

在所给数据中，存在的最大问题在于机械指标（TSS）、机械指标（色差）与人工指标中测试的黄桃样本互相不对应，不可直接用于拟合分析。我们不妨假设测试黄桃样本的指标在一定程度上代表了该批黄桃在该次测试中的状态情况，将该次测试的均值作为该批黄桃当前状态的指标。

通过批平均操作之后，我们可以将每批数据的机械指标（TSS）、机械指标（色差）与人工指标对应起来，用于后续的拟合、分类、预测等任务。

2.2.6 关于不同年份记录数据不同的处理说明

在所给数据中，2018 年机械指标只提供了带皮硬度、去皮硬度、TSS 和色差指标，而 2019 年数据提供了包括桃左右侧的带皮硬度、去皮硬度、TSS 和色差指标，以及增加了果实大小（纵经、横经、侧经）等数据，2020 年的数据中又舍弃了果实大小数据。为了统一这些数据便于我们的模型构建、处理和分析，我们将果实大小数据舍弃、并对 2019 年桃的左右侧机械指标取平均作为该样本的机械指标。

3 模型建立与求解

3.1 问题分析

问题一有三个要求，分别是比较指标评分之间的关系、机械指标与人工指标评判一致性和构建黄桃品质分析的定量模型并用该模型评判“专家经验”和“普遍认知”。

机械指标和人工指标关系可以用两两指标之间的散点图，指标本身的分布直方图来评估，判断不同品质的桃子我们可以用平行坐标图来看中位数、四分位数变化，进一步分析关系我们可以使用主成分分析法（PCA）来在主成分空间中观测不同指标的相关性和指标体系中的冗余度，由于各分量动态范围不同，使用的主成分分析方法应该是加权的（WPCA）。根据题目背景，指标可以分为三类，一是人工指标，二是机械主要指标，三是机械色差指标。在各指标自身分析时我们以单个黄桃为最小研究对象，在分析两类指标关系时我们以每一批黄桃为研究对象。由于 2019 人工打分素质较高，并且色差与机械指标一一对应，我们用 2019 年数据来进行指标间关系分析。

机械指标体系和人工评分体系，都是用了三个维度的指标，为了评判两个体系的一致性，我们应该建立两者的“映射”，2018 年和 2019 年每一批数据都同时在两个体系中进行了打分，因此，通过这些数据我们可以完成两个体系之间的映射，映射方法分为线性方法（如线性回归模型），非线性方法（如神经网络）。我们建立好两个体系的映射后，在机械评分体系中根据桃子的分布（多维正态分布）随机生成大量测试点直接进行机械打分，然后映射至人工评分体系用人工评分要求进行优劣判别，这样就可以得到同一个测试点在两个体系中的评价从而进行一致性分析。

当我们实现了上述两个评分体系的映射后，我们可以直接提出一个严格的品质定量模型，即为两个体系交叉部分为优，其余为劣。但是以上的体系都只用到了部分指标，而其余指标（如芳香、桃子颜色）的确是黄桃品质的重要判断依据。

为了利用尽可能多的指标，以及衡量不同指标的重要性程度，我们采用层次分析法进行模型构建。最后，映射到同一个评价空间上也可以方便地判断机械评分和人工评分的包容关系也就是分析“专家经验”和“普遍认知”。

3.2 指标间关系

(1) 人工指标关系分析

做出四类人工指标的散点图如下：

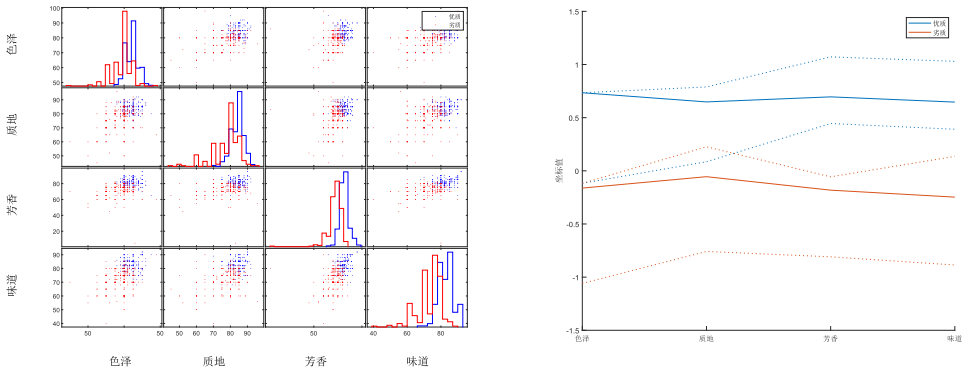


图 5 人工指标散点图（左）及人工指标平行坐标图（右，实线为中位数线，虚线为四分位数线）

通过左图我们可以看到，四个指标之间是具有相关性的，通过右图人工评分的优劣跟芳香也很有关系，因此在后面的层次分析体系中加入芳香是很有必要的。进一步，我们用加权主成分分析（WPCA）分析指标间关系：

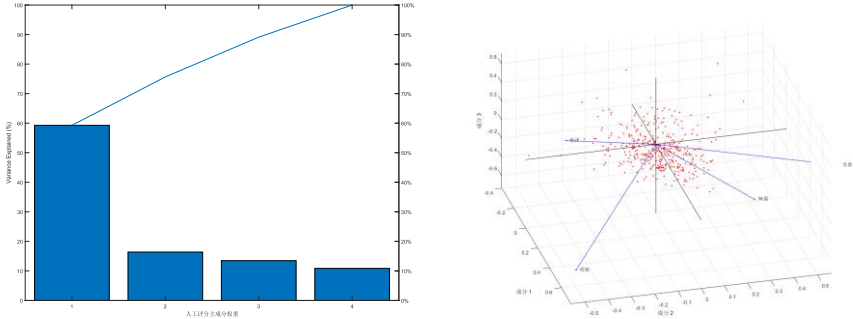


图 6 人工指标 WPCA 权重图（左）及指标关系图（右）

通过左图能量占比我们可以发现，四个指标的关联度其实很高，一到两个主成分就具有绝大部分信息。通过右图我们可以看到芳香和味道关系密切。

(2) 机械指标关系分析

我们先看色差的指标：

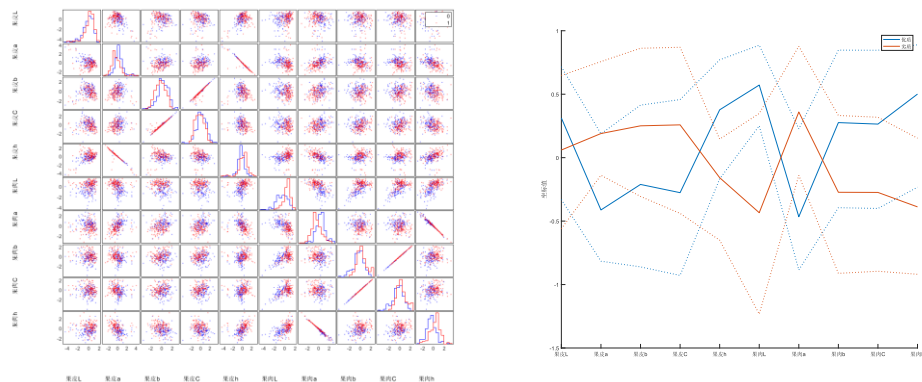


图 7 色差指标相关性（左）及不同品质黄桃样品色差分布图

通过左图我们可以明显看到色差的十个指标具有强相关性，通过右图我们可以看到优质和劣质桃子在不同色差范围内也是分布不同的，因此在层次分析体系中加入色差是合理的。

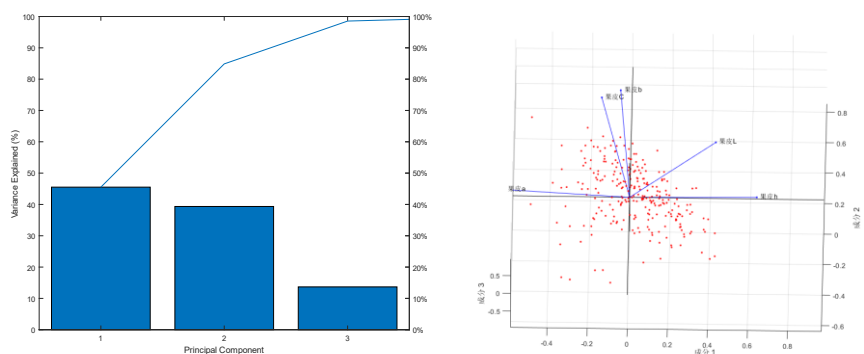


图 8 色差指标 WPC 权重图（左）及色差指标关系图（右）

我们取出果皮的五个色差指标进行 WPCA 分析，通过左图可以看到三个主成分即可包含所有信息，右图的指标轴关系也充分说明了这一点。后来查阅资料发现，色差的 L、a、b、c、H 五个分量具有以下关系：

$$\begin{aligned} a^2 + b^2 &= c^2 \\ \tan(H^\circ) &= b/a \end{aligned} \quad (1)$$

因此后面的分析中我们只用 L、a、b 三个分量来表示色差信息。再看机械主要指标

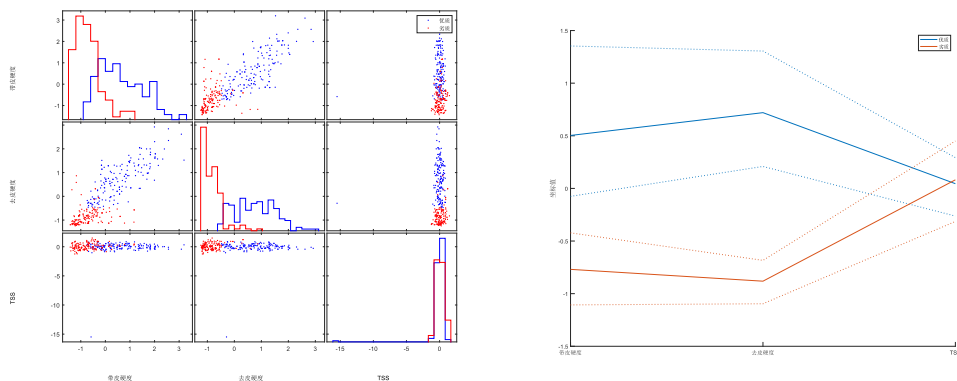


图 9 机械指标相关性（左）及不同品质黄桃样品机械指标分布图

通过左图我们可以明显看到带皮硬度和去皮硬度有明显相关性，而两者与 TSS 下降率却呈现不相关。右图可以看到，在 2019 年所有桃子中，无论是优桃还是劣桃，TSS 都比较小，对品质评判影响不大。

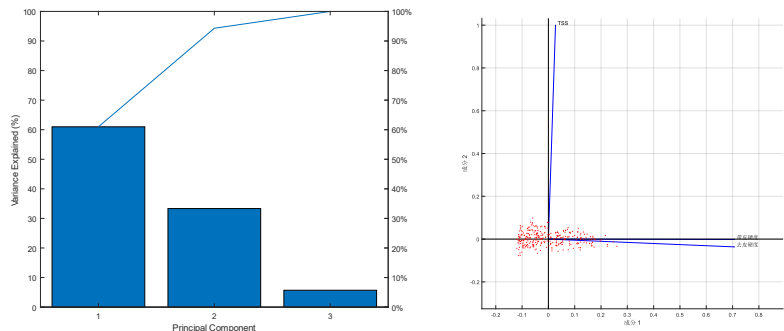


图 10 机械指标 WPCA 权重图（左）及机械指标相关性（右）

Wpca 的结果印证了这一点，两个主成分就可以解释绝大部分信息，去皮硬度和带皮硬度具有强关联性。

（3）机械指标与人工指标关系分析

先看人工评分与机械指标

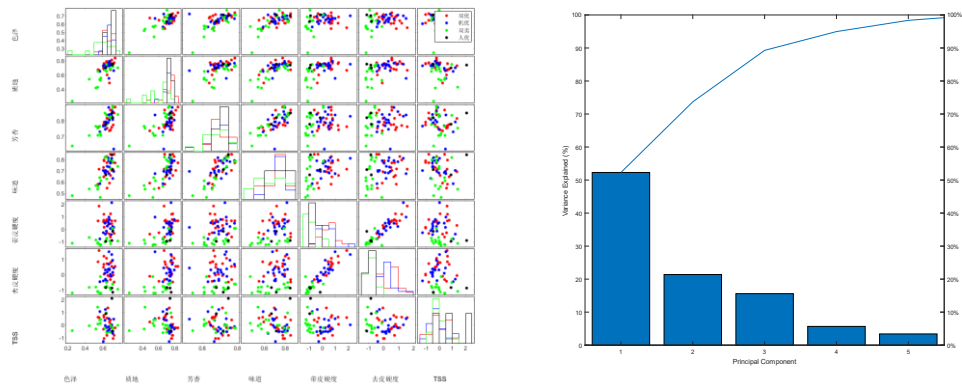


图 11 人工指标与硬度相关性（左）及 WPCA 权重图（右）

通过左图可以看到带皮（去皮）硬度与质地具有较强相关性，TSS 与味道具有一定相关性，在 WPCA 分析中，三个维度足以解释七个指标所有信息，六个指标具有很强的相关性。

再看人工评分和色差指标

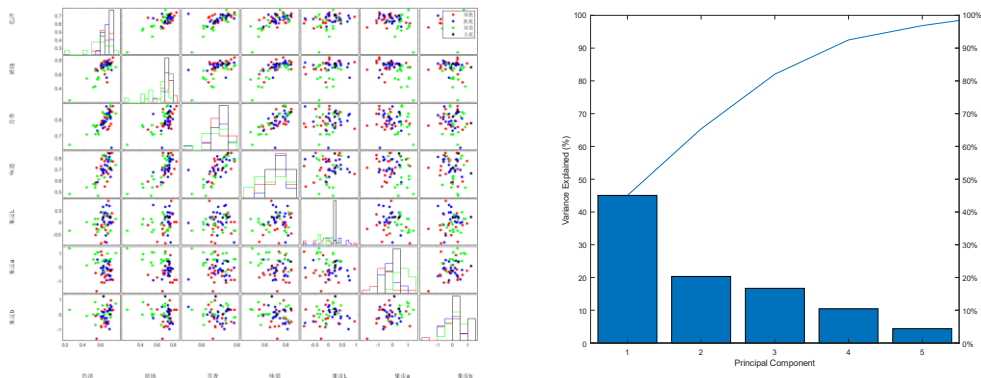


图 12 人工指标与色差相关性（左）及 WPCA 权重图（右）

通过左图可以看到味道色泽跟 L 具有一定相关性，在 WPCA 分析中，需要

四个维度来解释绝大部分信息，七个指标的相关性不大。

(4) 两类指标相关性结论

通过 WPCA 分析，取前五个主成分构成线性空间，通过主成分权重系数计算可得各个指标相关性系数如下：

	色泽	质地	芳馨	味道	带皮硬度	去皮硬度	TSS	果皮 L'	果皮 a'	果皮 b'	果肉 L'	果肉 a'	果肉 b'
色泽	1	0.707329	0.906291	0.790741	-0.18866	-0.16087	0.067986	0.380604	-0.10728	0.52771	0.233282	-0.23611	0.2189828
质地	0.707329	1	0.957611	1.035656	0.378024	0.21571	-0.39717	0.313869	-0.01606	-0.31447	0.226782	0.100662	-0.13083
芳馨	0.906291	0.957611	1	0.866383	-0.14548	-0.23203	0.212864	-0.2941	0.104444	-0.04623	-0.03201	-0.11162	-0.137366
味道	0.790741	1.035656	0.866383	1	0.123875	-0.02415	-0.19709	-0.21751	0.118264	-0.24972	0.062534	0.071095	-0.084992
带皮硬度	-0.18866	0.378024	-0.14548	0.123875	1	0.889995	-0.42717	0.153781	-0.42441	-0.93485	0.44866	-0.07045	0.0304854
去皮硬度	-0.16087	0.21571	-0.23203	-0.02415	0.889995	1	-0.32968	0.249008	-0.65522	-0.7285	0.642051	-0.30968	0.3194907
TSS	0.067986	-0.39717	0.212864	-0.19709	-0.42717	-0.32968	1	-0.05678	-0.23961	-0.05031	-0.05522	-0.55091	-0.331602
果皮 L'	0.380604	0.313869	-0.2941	-0.21751	0.153781	0.249008	-0.05678	1	-0.09247	0.197383	0.073965	0.0449	-0.253276
果皮 a'	-0.10728	-0.01606	0.104444	0.118264	-0.42441	-0.65522	-0.23961	-0.09247	1	0.052738	-0.75366	1.036182	-0.517185
果皮 b'	0.52771	-0.31447	-0.04623	-0.24972	-0.93485	-0.7285	-0.05031	0.197383	0.052738	1	0.061316	-0.05721	0.5113481
果肉 L'	0.233282	0.226782	-0.03201	0.062534	0.44866	0.642051	-0.05522	0.073965	-0.75366	0.061316	1	-0.7641	1.125246
果肉 a'	-0.23611	0.100662	-0.11162	0.071095	-0.07045	-0.30968	-0.55091	0.0449	1.036182	-0.05721	-0.7641	1	-0.293679
果肉 b'	0.218982841	-0.130830405	-0.137365509	-0.084992278	0.030485383	0.319490736	-0.331602402	-0.253275646	-0.517184781	0.511348133	1.123245963	-0.293678728	1

图 13 各指标相关性系数

3.3 基于两种回归算法的评判一致性分析

为了分析机械指标与人工评分之间的关系以及两种方法的评判一致性，显然，我们需要将机械指标与人工评分映射到同一个域中。由于人工评分更加地直观，我们不妨将机械指标映射到人工指标域中，然后利用 18 年和 19 年的机械与人工评分数据进行回归分析，从而对两种方法进行一致性评判。

为了保持 18 年和 19 年数据的延续性，对机械指标评分我们采用带皮硬度 $> 1.5\text{kg}/\text{cm}^2$ 、去皮硬度 $> 1.5\text{kg}/\text{cm}^2$ 和 TSS 增长率 $> -20\%$ 作为优质判断标准。对于人工指标，我们采用色泽、质地和味道均分大于 80 作为优质判断标准。

为了将机械指标映射到人工指标域，且防止由于数据太少造成数据过拟合或者欠拟合，我们提出了两种映射方法：a. 基于反向传播算法的线性回归，b. 基于神经网络的非线性回归。通过这两种方法的彼此比较和分析，证明了我们所提方法的有效性。

由于 19 年的数据记录较为规范严格，我们先用 19 年的数据对两种评判方法进行关系拟合和一致性分析。

3.3.1 基于反向传播算法的线性回归

在一般认知中，一个桃子的机械指标评分与其人工评分存在正相关，我们不妨假设机械指标与人工评分线性相关，即机械指标可以通过线性映射投影至人工评分域。

$$Score_{man} = F(Score_{mec}) \quad (2)$$

其中， $Score_{man} \in R^{n \times 3}$ 代表人工标准， $Score_{mec} \in R^{n \times 3}$ 代表机械指标标准， n 代表用于拟合的数据批数， $F(\cdot)$ 代表线性映射函数，设 $k \in R^{3 \times 3}$ 为映射系数， $b \in R^{1 \times 3}$ 为偏置， $F(\cdot)$ 可展开为

$$F(x) = kx + b \quad (3)$$

对于多维线性映射来说，直接通过公式求解比较复杂，于是我们采用反向传

播算法对该模型进行求解：以 $Score_{mec}$ 为输入， $Score_{man}$ 为输出，定义损失函数为

$$MSE = \frac{1}{n} \sum_n ||Score_{predman} - Score_{man}||^2 \quad (4)$$

通过使用学习率 0.001 的 Adam 优化器更新参数，直至 MSE 不再变化。以 2019 年的数据为例，最终 $MSE = 0.4386$ ，映射系数 k 和偏置 b 如下：

$$k = \begin{bmatrix} 0.208 & 0.557 & 0.358 \\ 0.110 & -0.06 & -0.06 \\ -0.113 & -0.19 & -0.159 \end{bmatrix} \quad (5)$$

$$b = [-0.027 \quad -0.032 \quad -0.025] \quad (6)$$

为了检验拟合效果，我们利用原始数据的均值与协方差生成符合多维正态分布的 5000 个机械指标评分数据，如下图所示：

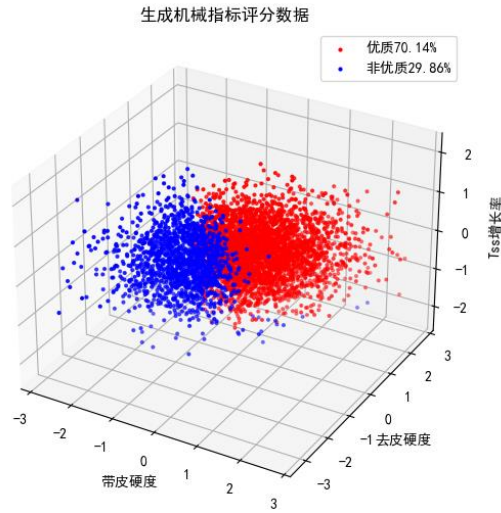


图 14 机械指标评分数据

利用线性映射函数，我们将这些数据映射到人工评分域，并画出人工评分域中利用机械指标评分和人工评分的结果如下：

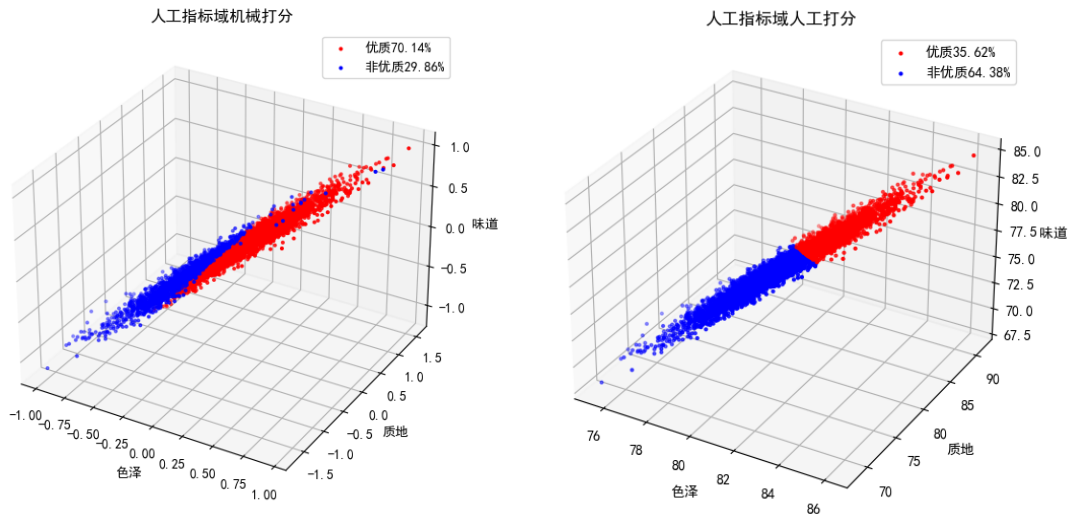


图 15 人工评分域评分结果

可以看出，该线性映射方法的结果较好，人工评分的结果比机械指标评分的结果更为严格且基本包含在机械评分的结果之中，符合题目中对于 2019 年人工评分较为严格的设定。为了进一步证明以上结论，我们将人工评分的结果映射回机械评分域对比如下：

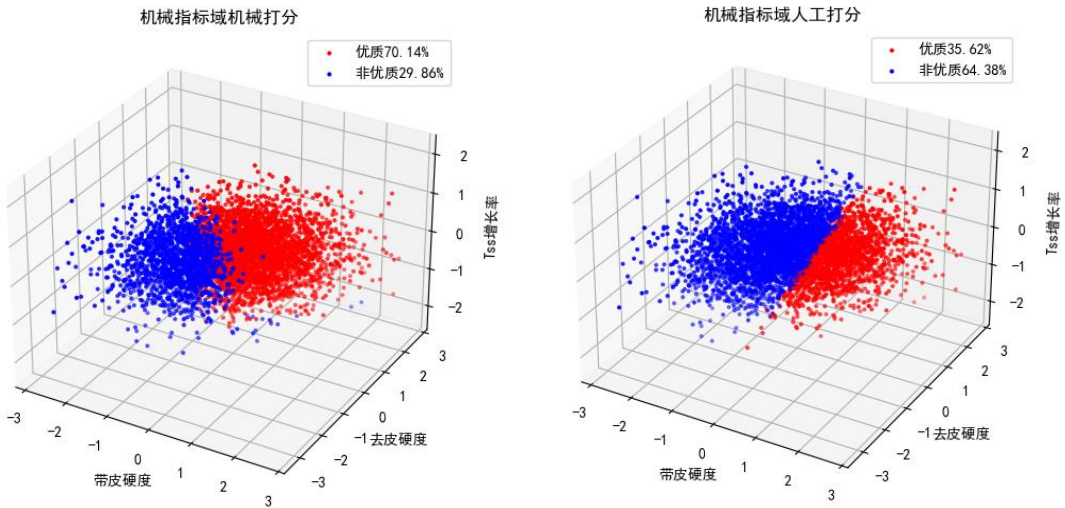


图 16 机械指标域评分结果

证明模型的有效性之后，进一步我们利用该模型对机械指标和人工评分进行一致性分析。利用生成的 5000 个数据和线性映射模型，生成两种方法评判结果的混淆矩阵和维恩图如下图所示：

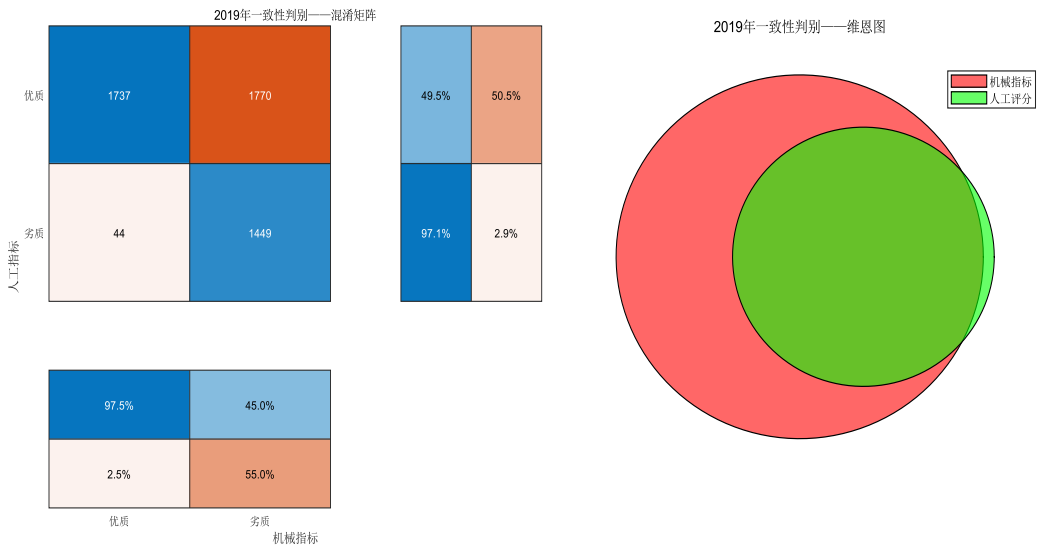


图 17 2019 年一致性判别混淆矩阵和维恩图（线性回归）

从图中可以看出，在本批数据中心，人工指标与机械指标有较大的重合度，人工指标判断为优质的，机械指标有较大概率（97.5%）判断为优质；而反过来机械指标判断为优质的，人工指标只有 49.5%的概率同样判断为优质，说明人工指标要更为严格。为了更直观地看出人工指标和机械指标对这批数据的评判一致性，我们画出 3D 散点图如下：

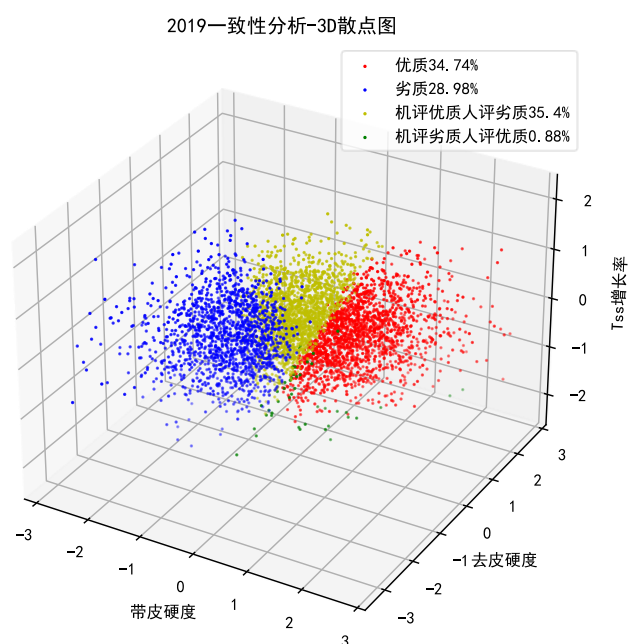


图 18 2019 年一致性分析（线性回归）

从图中我们可以看出，两种评判方法都觉得优质的概率有 34.74%，都觉得劣质的概率有 28.98%，也就是说，两种方法做出同样判断结果的概率有 63.72%。而机械指标评分优质而人工评分劣质的数据数量占了总数据的 35.4%，再次反映了题中提到的人工指标较为严格。

因此，机械指标和人工评分方法的在满足“人工评分较为严格”这一条件下具有较好的一致性。

为了定量描述这两种方法的一致性，我们引入精确率（Precision）、召回率（Recall）、准确率（Accuracy）、F1-Score、Kappa 系数等评价指标，他们的定义和计算结果如下表所示：

表 3 评价指标定义及结果

指标	定义	数值
TP	机械与人工评分都为优质	1737
TN	机械与人工评分都为劣质	1449
FP	机械评分为优质，人工评分为劣质	1770
FN	机械评分为劣质，人工评分为优质	44
Precision	$\frac{TP}{TP + FP}$	0.4953
Recall	$\frac{TP}{TP + FN}$	0.9753
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	0.6372
F1-Score	$\frac{2PR}{P + R}$	0.6570
P_e	$\frac{(TP + TN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}$	0.4421
Kappa	$\frac{OA - P_e}{1 - P_e}$	0.3497

从表格中我们可以看出，两种评判方法的结果具有较高的召回率（0.9753）和精度（0.6372），召回率高说明机械指标能够很好地预测人工评分为优的黄桃，而精度反映了两种评价方法的预测一致性。

此外，我们发现两种评判方法的 Kappa 系数为 0.3497，根据下图所示的 Kappa 系数变化与一致性之间的关系，我们发现按照 Kappa 系数进行评判，这两种方法的一致性评判结果为一般。

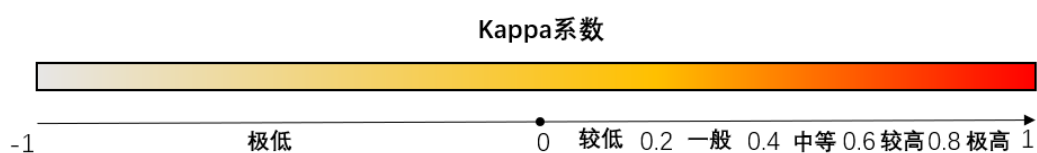


图 19 Kappa 系数与一致性关系

3.3.2 基于神经网络的非线性回归

为了增加拟合模型的非线性性，使得模型能够更好地拟合数据，我们提出了一个基于神经网络的非线性模型，如下图所示：

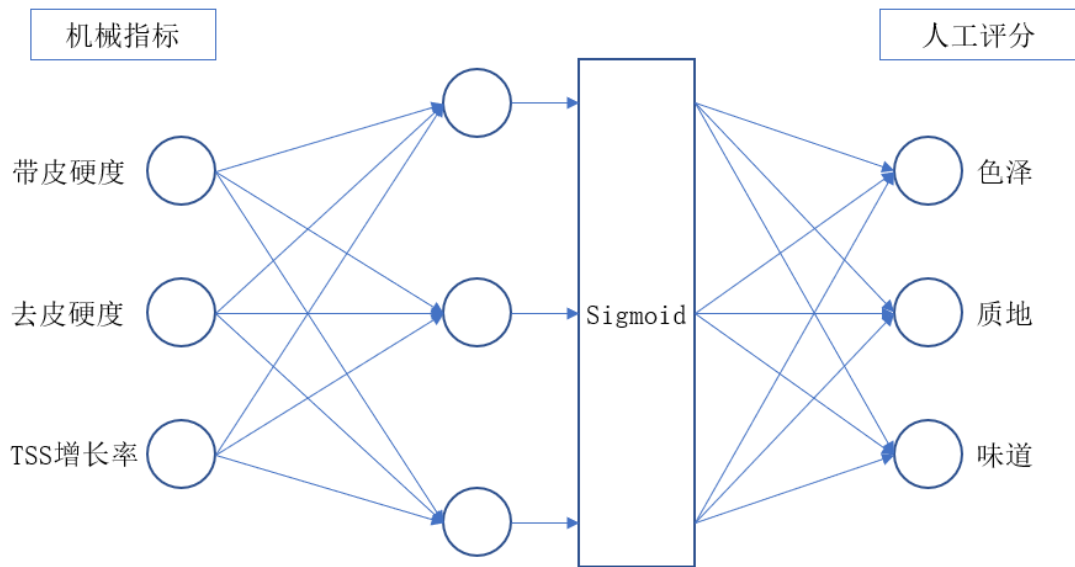


图 20 神经网络非线性模型图

为了防止网络过拟合,我们采用只有一个隐藏层(三个隐藏神经元)的结构。使用 sigmoid 使得模型有了非线性表达能力。以 MSE 作为损失函数,使用学习率 0.001 的 Adam 优化器训练模型。

同样,我们先利用多维正态分布生成 5000 个样本的黄桃机械评分数据,如下图所示:

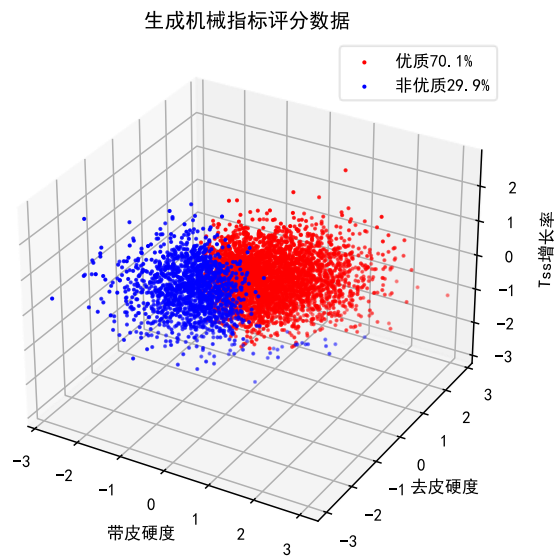


图 21 机械指标评分

接着使用训练好的神经网络模型将机械指标映射到人工指标域,人工指标域的机械打分图和人工打分图如下图所示:

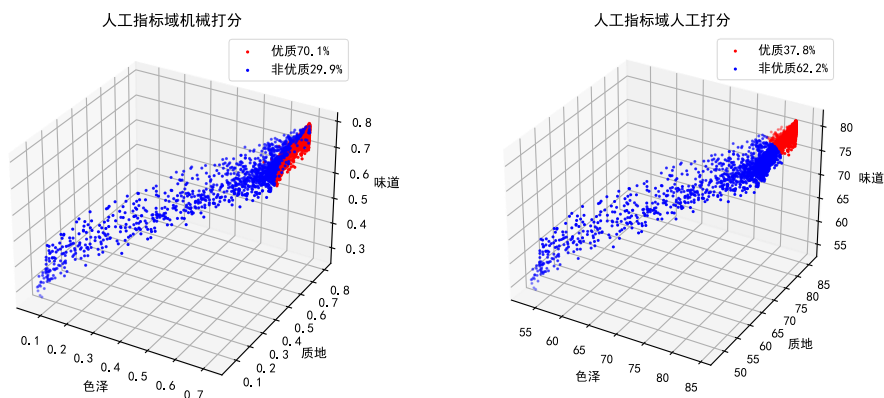


图 22 人工指标域评分

可以看出，神经网络模型对于机械指标和人工评分的映射具有非线性性，倾向于将机械指标映射到评分 70~90 的区间内，更符合实际人工评分数据的分布情况。

进一步，将人工评分结果映射回机械指标域结果如下图所示：

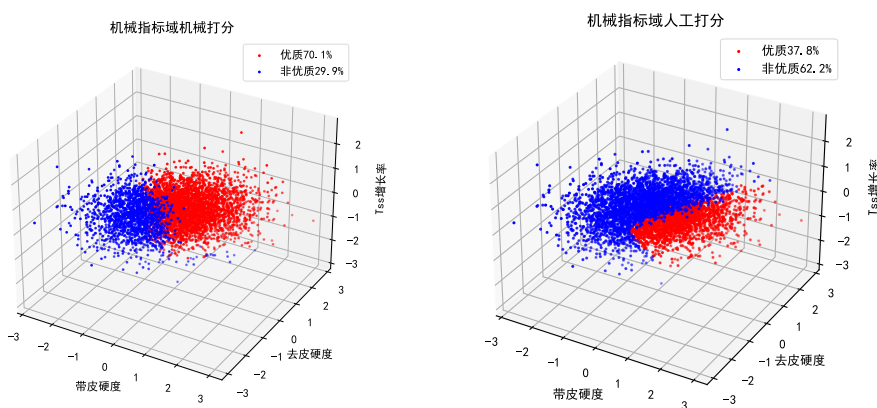


图 23 机械指标域评分

从图中可以看出有部分数据虽然机械评分高，但是人工评分低。通过我们对两种贮藏方式和测试时间的统计观察可以发现，在冷库 1° C 中贮存的黄桃存在硬度变大而人工评分变低的情况，上图正好拟合了这种情况。上面的人工评分域和机械指标域的映射图都证明了我们使用神经网络进行非线性映射的效果比较好。

利用神经网络拟合结果我们对两项指标进行一致性分析，画出其混淆矩阵图、维恩图、交叉散点图如下所示：

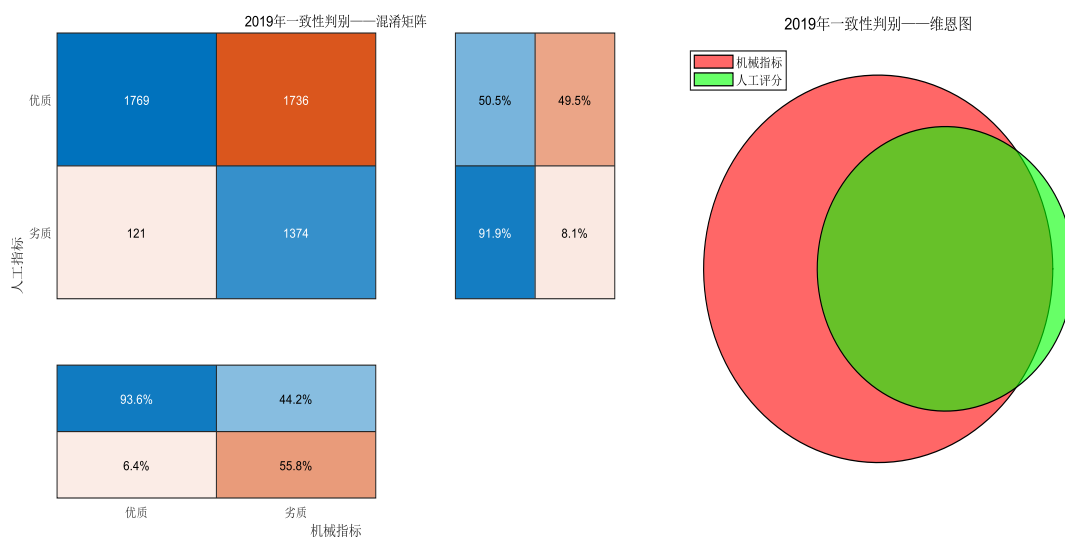


图 24 2019 年一致性判别混淆矩阵和维恩图（非线性回归）

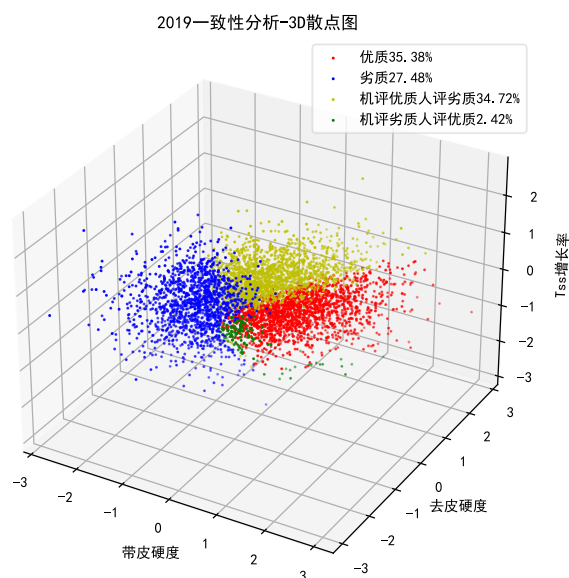


图 25 2019 年一致性分析（非线性回归）

可以看出，两种方法同评优质的概率为 35.58%，同评劣质的概率为 27.48%，因此两种做出一致评判的概率为 62.86%，且人工评分为优的黄桃机械指标为优的概率为 93.6%，再次印证了前提条件“人工指标较为严格”。

进一步，我们利用精确率（Precision）、召回率（Recall）、准确率（Accuracy）、F1-Score、Kappa 系数等评价指标，他们的定义和计算结果如下表所示：

表 4 评价指标定义及结果

指标	定义	数值
TP	机械与人工评分都为优质	1769
TN	机械与人工评分都为劣质	1374
FP	机械评分为优质，人工评分为劣质	1736
FN	机械评分为劣质，人工评分为优质	121
Precision	$\frac{TP}{TP + FP}$	0.5047
Recall	$\frac{TP}{TP + FN}$	0.9360
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	0.6286
F1-Score	$\frac{2PR}{P + R}$	0.6558
P_e	$\frac{(TP + TN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}$	0.4510
Kappa	$\frac{OA - P_e}{1 - P_e}$	0.3235

可以看出，通过非线性神经网络对两种评判指标映射关系拟合后的定量一致性分析结果与采用线性回归拟合后得到的定量一致性分析结果较为接近，保持着高召回率（0.9360）和较高准确率（0.6286）。Kappa 系数为 0.3235，结论依然是两种评判方法的一致性一般，符合“人工评分较为严格”这一前提。

此外，我们还对 18 年的数据进行了拟合与一致性分析。使用线性回归算法拟合，得到的映射系数 k 和偏置 b 如下：

$$k = \begin{bmatrix} -0.834 & -0.233 & -0.617 \\ 1.035 & 0.120 & 0.260 \\ 0.118 & 0.053 & 0.176 \end{bmatrix} \quad (7)$$

$$b = [0.022 \quad 0.024 \quad 0.025] \quad (8)$$

映射的结果如图所示：

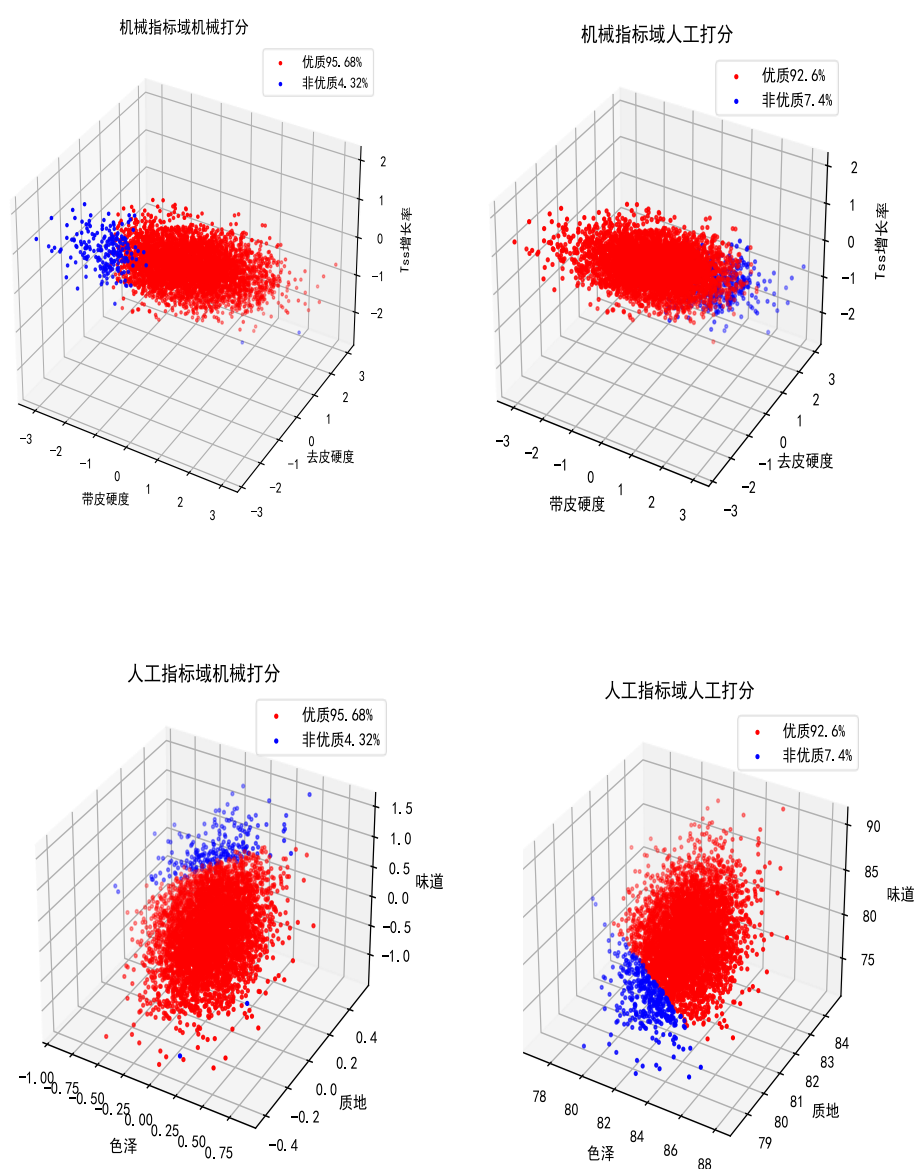


图 26 2018 年人工指标域和机械指标域评分图

通过上图我们可以看出，对 18 年做线性回归的分析显示，机械打分为优质的概率为 95.68%，而映射到人工指标后人工评分为优质的概率为 92.6%。对比 18 年的实际数据，机械打分为优质的概率为 90%，证明我们的多维正态分布假设较好地拟合实际分布。对比 19 年的数据，我们可以发现 18 年的机械打分人工打分为优的概率都比较大，一方面是由于我们生成数据的分布拟合的是 18 年实际数据分布，另一方面是由于 18 年人工评分没有 19 年严格。

利用拟合后的模型对 18 年的两种方法进行一致性分析，首先画出混淆矩阵、维恩图和散点图如下：

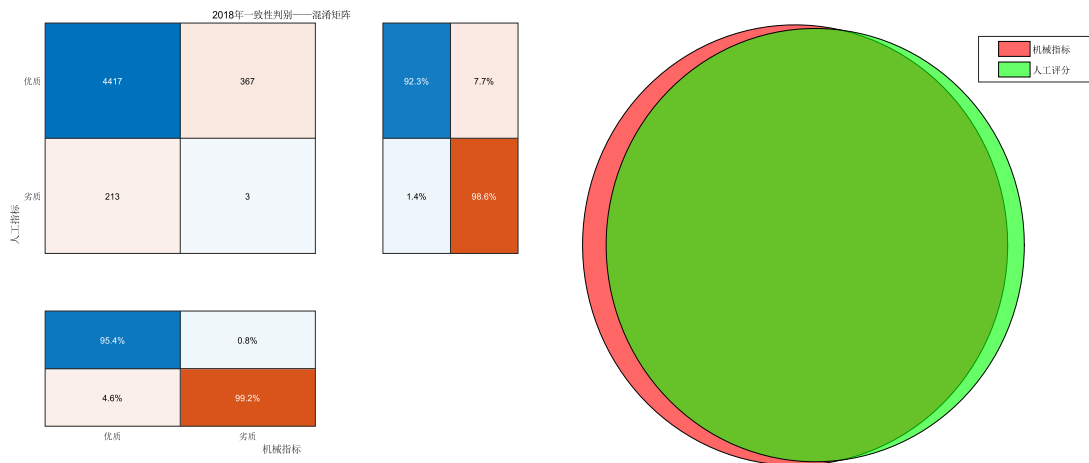


图 27 2018 年一致性判别混淆矩阵和维恩图

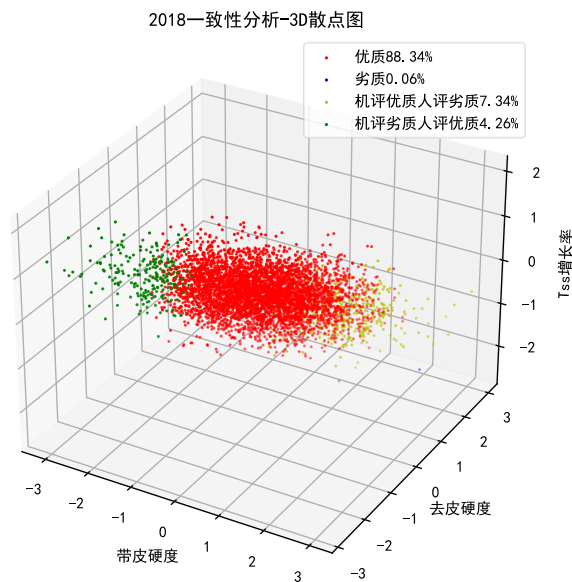


图 28 2018 年一致性分析

通过上图可以看出，人工评分与机械评分同优的概率为 88.34%，看似很高，其实是因为根据 18 年分布生成的数据中优秀的比例本身就很高。而将人工指标域人工打分和机械打分的两个图对比可以看出，两种方法对非优质数据的评判其实并不一致。因此基于 18 年的数据，两种方法的评判一致性并不好。

为了证明我们的观点，我们利用精确率（Precision）、召回率（Recall）、准确率（Accuracy）、F1-Score、Kappa 系数等评价指标，他们的定义和计算结果如下表所示：

表 5 评价指标定义及结果

指标	定义	数值
TP	机械与人工评分都为优质	4417
TN	机械与人工评分都为劣质	3
FP	机械评分为优质，人工评分为劣质	367
FN	机械评分为劣质，人工评分为优质	213
Precision	$\frac{TP}{TP + FP}$	0.9233
Recall	$\frac{TP}{TP + FN}$	0.9540
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	0.8840
F1-Score	$\frac{2PR}{P + R}$	0.9384
P_e	$\frac{(TP + TN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}$	0.8892
Kappa	$\frac{OA - P_e}{1 - P_e}$	-0.0469

可以看出，虽然其他各项指标都比较高，但是 Kappa 系数却只有 -0.0469，说明其他指标高只是因为 TP 高，即由于符合优质指标的数据较多导致的机械评分和人工评分都评判为优质的比例较高，拉高了与 TP 有关的评价指标。但 Kappa 系数缺很低，证明基于 18 年的数据，两种方法的评判一致性确实并不好。

3.4 黄桃品质分析的定量模型

为了为黄桃品质构建定量分析模型，我们采取了两种思路。一种是从果实的机械指标（带皮硬度、去皮硬度、TSS 增长率）测量数据出发，利用 3.3 中使用基于 2019 年数据训练出的回归模型，基于 2019 年数据“人工指标比机械指标更为严格，且人工评优有大概率被机械评优”这一原则，利用机械指标映射得到的人工指标评分作为定量分析结果，这是一种简单评判模型，但其严格且有效。第二种是将色差数据加入到机械指标之中，利用层次分析法综合分析包括机械指标和人工指标在内的各项指标的重要性，最终给出综合评分来判断黄桃的优质与否，这是一种考量各指标后得出的综合评判模型。

3.4.1 基于回归分析的评价模型

利用 3.3 中提到的回归方法，利用测试得到的机械指标数据，我们便可以通过线性或非线性方法映射为人工评分，为使得模型严格有效，我们取机械指标域映射人工指标共同为评价为优的结果做为评判结果。该模型的描述如下图所示：

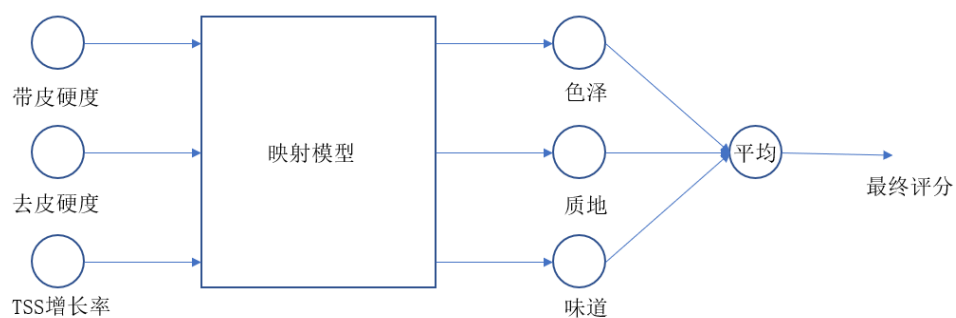


图 29 基于回归分析的评价模型

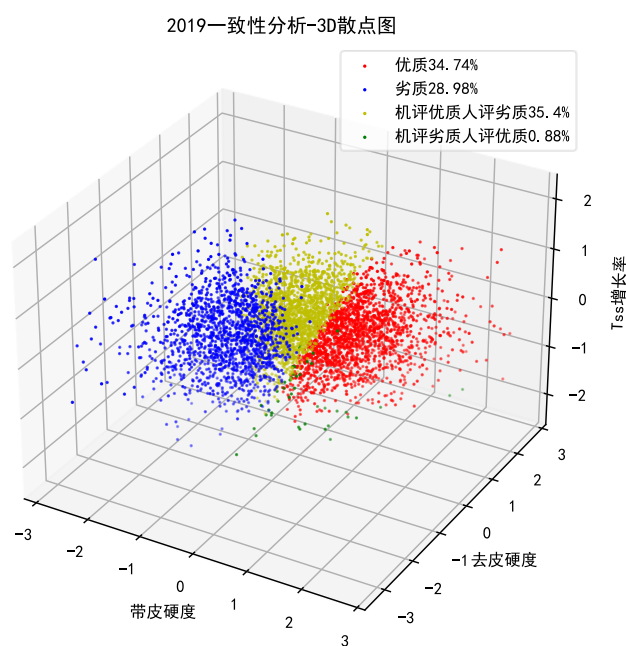


图 30 基于评价模型得出的 2019 年一致性分析

利用该模型得到的结果,在 2019 年数据上的拟合率(判断一致性)为 73.21%,能够较好地拟合实际评分结果。

3.4.2 基于层次分析法的评价模型

3.4.2.1 根据色差进行分类分析

根据潘通色卡将色差仪检测得到的 L (明暗度)、a (红绿色)、b (黄蓝色) 值进行分类分析,将黄桃颜色分为黄色、浅黄色和暗黄色^[1]。

表 6 颜色标准 Lab 值 PANTONENC 对照表^[2]

序号	PANTONENC	L	a	b	色样
1	100C	93	-8	50	
2	101C	92	-9	66	
3	102C	91	-7	100	
4	103C	71	0	91	
5	104C	62	-1	74	
6	105C	48	-1	49	
7	106C	91	-6	65	
8	107C	90	-4	81	
9	108C	88	0	95	
10	109C	86	5	100	
11	110C	73	7	94	
12	111C	61	4	75	
13	112C	55	2	62	
14	113C	90	-2	67	
15	114C	89	-2	70	
16	115C	89	-1	72	
17	116C	85	7	93	
18	117C	68	10	83	
19	118C	60	8	73	
20	119C	50	1	49	
21	120C	89	0	53	
22	1205C	90	0	31	
23	121C	89	2	57	
24	1215C	89	1	40	
25	1225C	87	5	68	
26	123C	84	11	69	
27	1235C	84	12	79	
28	124C	79	22	85	
29	1245C	75	16	85	
30	125C	64	13	69	
31	1255C	61	12	72	
32	1265C	56	10	59	
33	127C	55	8	60	
34	128C	46	7	46	
35	129C	89	-3	46	
36	130C	86	0	61	
37	131C	84	3	69	

设立四种色样集 U_1 、 U_2 、 U_3 分别代表黄色黄桃色样集、浅黄色黄桃色样集和暗黄色黄桃色样集，以表 2 中的色样序号为集合中的元素。通过目视比色法，则可以得到：

$$U_1 = \{2,3,7,8,9,10,14,15,16,17,21,23,25,26,27,36,37\}$$

$$U_2 = \{1,22,24,35\}$$

$$U_3 = \{4,5,6,11,12,13,18,19,20,28,29,30,31,32,33,34\}$$

为了衡量不同色样集之间的界限，我们对色样集元素的 Lab 值进行多重比较。通过比较色样集元素的 L 值，我们发现黄色和浅黄色色样集元素的 L 值没有显著区别，但是暗黄色集元素的 L 值与这两个色样集差异明显。因此我们可以通过 L 值来区分暗黄色黄桃与其他两种颜色的黄桃。

表 7 色样集元素 L 值差异表

色样集	L 值
U_1, U_2	[80,93]
U_3	[46,80)

再比较色样集元素的 a、b 值，我们发现黄色色样集和浅黄色色样集元素的 a 值无显著区别，但可以通过 b 值来区分黄色黄桃和浅黄色黄桃。

表 8 色样集元素 a、b 值差异表

色样集	a 值	b 值
U_1	[-9,12]	(50,100]
U_2	[-8,3]	[40,50]

最终我们可以通过判别 L 值和 b 值来将黄桃样品归类到黄色黄桃、浅黄色黄桃或者暗黄色黄桃中。具体流程如下图所示：

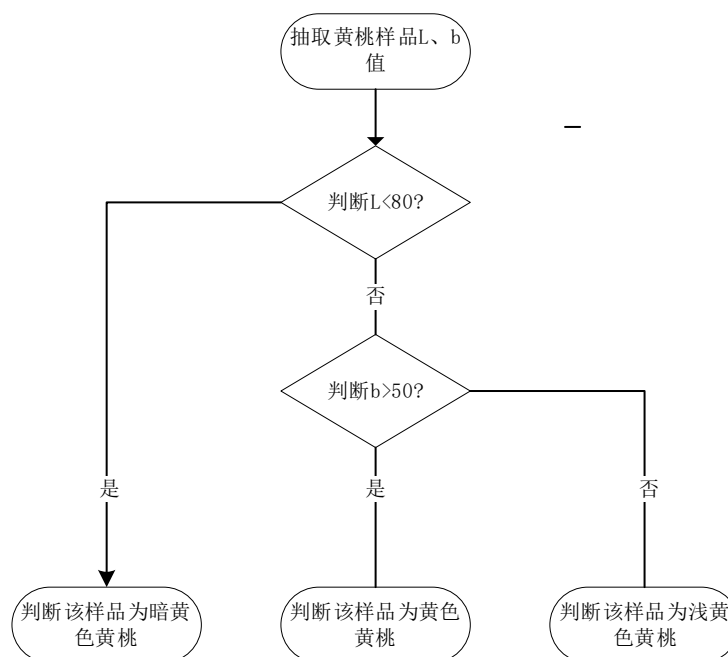


图 31 判断黄桃样品颜色流程图

3.4.2.2 模型建立

为了科学考量机械指标和人工评分，我们建立机械指标和人工评分的综合双评分机制，并且将黄桃颜色、人工评价的黄桃芳香分数也加入评价体系中。

对于人工评价我们采用百分制总分法，我们对色泽、质地、芳香和味道四个指标进行评价得出人工评价的最终得分 S。对于机械指标，我们对硬度、TSS、黄桃颜色三个指标进行评判得出机械指标得分 M，其中硬度取带皮硬度和去皮硬度中的最小值。

$$S = ax_1 + bx_2 + cx_3 + dx_4, 0 \leq a, b, c, d \leq 1, a + b + c + d = 1 \quad (9)$$

$$M = y_1 \oplus y_2 \oplus y_3, y_1, y_2, y_3 \in \{0, 1\} \quad (10)$$

$$F = M + \frac{S}{100} \quad (11)$$

品质 Q 的评定标准:

$$Q = \begin{cases} 1 (\text{优质}), & \text{if } F > 1 + \alpha \text{ or } F \in [0.9, 1], y_1 > 1.5 \text{kg/cm}^2, y_2 > 9\% \\ 0 (\text{劣质}), & \text{else} \end{cases} \quad (12)$$

其中 F 代表机械指标和人工评价的综合得分, α 为人工评价在综合得分中的权重因子。 $\alpha \in [0.75, 0.85]$, 若权重因子不在这个范围内, 令 $\alpha = 0.8$ 。 x_1, x_2, x_3, x_4 代表色泽、质地、芳香和味道四个影响因子, a, b, c, d 为四个影响因子的权重。 y_1, y_2, y_3 代表硬度、TSS 和黄桃颜色的判断因子。考虑到机械测量的误差以及人的口感体验, 我们设立当人的口感体验极佳时 (即当 $M = 0, S \geq 90$ 时), 且硬度和 TSS 满足设定最低限度, 就认为该黄桃品质是优质的。为了保证黄桃的食品标准, 通过查阅资料^{[3][4][5][6]}, 我们设定了硬度(y_1)和 TSS(y_2)的最低限度。

$$y_1 = \begin{cases} 1, & \text{if 硬度大于 } 1.5 \text{kg/cm}^2 \\ 0, & \text{else} \end{cases} \quad (13)$$

$$y_2 = \begin{cases} 1, & \text{if TSS 下降率小于 } 20\% \\ 0, & \text{else} \end{cases} \quad (14)$$

$$y_3 = \begin{cases} 1, & \text{if 黄桃颜色判定为黄色} \\ 0, & \text{else} \end{cases} \quad (15)$$

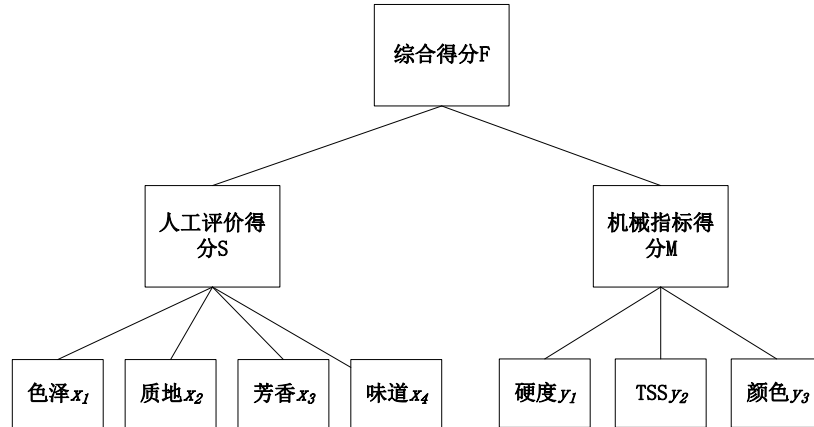


图 32 层次结构图

3.4.2.3 建立判断矩阵

设人工评价最终得出的分数为 S, 将色泽、质地、芳香和味道四个影响最终得分的因子设为 $X = \{x_1, x_2, x_3, x_4\}$ 。在确定某一个因子在影响最终得分中所占的

权重时，常常遇到的困难是难以将其影响程度量化。另外如果当影响因子数量较多时，直观去定论某一影响因子对最终结果有多大影响，往往会有失偏颇、顾此失彼，可能导致最终结果差强人意甚至数据出现矛盾。为了解决这种情况 Santy 等人提出了一种一致矩阵判断法。这种方法不是把所有影响因子放在一起比较，而是两两比较。比较时采用相对的尺度，以减少不同性质因子间比较的困难。

具体方法是：每次取出两个因子 x_i 和 x_j ($i, j \in \{1, 2, 3, 4\}, i \neq j$)，以 a_{ij} 表示 x_i 和 x_j 对 S 的影响大小之比，全部的比较结果用矩阵 $A = (a_{ij})_{4 \times 4}$ 表示，称 A 为 Z - X 的成对比较判断矩阵。

Santy 等人建议引用数字 1-9 及其倒数作为标度来确定矩阵中的元素 a_{ij} 的值。具体标度含义如下表所示：

表 9 判断矩阵中标度的含义

标度	含义
1	表示两个因素相比，具有相同重要性
3	表示两个因素相比，前者比后者稍重要
5	前者比后者明显重要
7	前者比后者强烈重要
9	前者比后者极端重要
2, 4, 6, 8	表示上述相邻判断的中间值
倒数	若因素 i 与因素 j 的重要性之比为 a_{ij} ，那么因素 j 与因素 i 重要性之比为 $a_{ji} = \frac{1}{a_{ij}}$

根据表 3，通过意愿调查，可以写出人工评分 S 的判断矩阵 A 如下：

S	x_1	x_2	x_3	x_4
x_1	1	1/2	1	1/3
x_2	2	1	1/2	1
x_3	1	2	1	1/2
x_4	3	1	2	1

同理，可以写出综合得分 F 的判断矩阵 A' 如下：

F	M	S
M	1	3
S	3	1

3.4.2.4 一致性检验

通过观察以上两个矩阵，可以发现矩阵存在以下两个性质：

$$1) a_{ij} > 0 \quad (16)$$

$$2) a_{ji} = \frac{1}{a_{ij}}, (i, j = 1, 2, \dots, n) \quad (17)$$

称满足以上两个条件的矩阵为正互反矩阵，对于正互反矩阵 A 有如下定义及定理：

定义 1：若矩阵 A 为正互反矩阵且满足 $a_{ij}a_{jk} = a_{ik} (i, j, k = 1, 2, \dots, n)$ ，则称该矩阵为一致矩阵。

定理 1：若 A 为一致矩阵，则

- 1) A 必为正反矩阵。
- 2) A 的转置矩阵 A^T 也是一致矩阵。
- 3) A 的任意两行成比例，比例因子大于零，从而 $\text{rank}(A) = 1$ 。
- 4) A 的最大特征值 $\lambda_{\max} = n$ ，其中 n 为矩阵 A 的阶。 A 的其余特征根均为零。

5) 若 A 的最大特征值 $\lambda_{\max}$ 对应的特征向量为 $W = (w_1, \dots, w_n)^T$ ，则 $a_{ij} = \frac{w_i}{w_j}, \forall i, j = 1, 2, \dots, n$ 。

若判断矩阵并不是一致矩阵，Santy 等人建议用其最大特征根对应的归一化特征向量作为权向量 W ，根据 $AW = \lambda W$ 可以计算得到 W 。该权向量的分量对应同一层次的相应因子对上一层次某因子的影响权重。

定义一致性指标：

$$CI = \frac{\lambda - n}{n - 1} \quad (18)$$

其中 n 为一致矩阵的阶，也是其唯一非零特征根， λ 为判断矩阵的最大特征根，当且仅当 $\lambda = n$ 时， A 为一致矩阵。可以用 $\lambda - n$ 的数值大小来衡量 A 不一致程度，数值越大，则不一致程度越大，引起的判读误差也会越大。

定义一致性比例：

$$CR = \frac{CI}{RI} \quad (19)$$

为了衡量 CI 的大小，Santy 等人引入了随机一致性指标 RI ， RI 值如下表所示：

表 10 判断矩阵中标度的含义

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

一般地，我们认为当一致性比例 $CR = \frac{CI}{RI} < 0.1$ 时，认为判断矩阵的一致性是可以接受的，否则应对判断矩阵作适当修正。

3.4.2.5 模型求解

通过一致性检验我们发现人工评分 S 的判断矩阵 A 不是一致矩阵，计算得出

矩阵 A 的最大特征值 $\lambda_{max} = 4.2699$ ，对应的归一化特征向量 $W = \{0.1495, 0.2402, 0.2480, 0.3623\}$ 。即对应的 $a = 0.1495, b = 0.2402, c = 0.2480, d = 0.3623$ 。

其一致性检验为：

$$CI = \frac{\lambda - n}{n - 1} = 8.9967 \times 10^{-2} \quad (20)$$

$$CR = \frac{CI}{RI} = 9.9963 \times 10^{-2} < 0.1 \quad (21)$$

则可以认为判断矩阵 A 的一致性是可以接受的。

综合得分 F 的判断矩阵 A' 是一致矩阵，计算的出其特征值 $\lambda = 2$ ，对应的归一化特征向量 $W' = \{0.25, 0.75\}$ 。即对应的 $\alpha = 0.75$ 。

则模型为：

$$S = 0.1495x_1 + 0.2402x_2 + 0.2480x_3 + 0.3623x_4, 0 \leq a, b, c, d \leq 1, a + b + c + d = 1 \quad (22)$$

$$M = y_1 \oplus y_2 \oplus y_3, y_1, y_2, y_3 \in \{0, 1\} \quad (23)$$

$$y_1 = \begin{cases} 1, & \text{if 硬度大于 } 1.5 \text{ kg/cm}^2 \\ 0, & \text{else} \end{cases} \quad (24)$$

$$y_2 = \begin{cases} 1, & \text{if TSS 下降率小于 } 20\% \\ 0, & \text{else} \end{cases} \quad (25)$$

$$y_3 = \begin{cases} 1, & \text{if 黄桃颜色判定为黄色} \\ 0, & \text{else} \end{cases} \quad (26)$$

$$F = M + \frac{S}{100} \quad (27)$$

品质 Q 的评定标准：

$$Q = \begin{cases} 1 (\text{优质}), & \text{if } F > 1.75 \text{ or } F \in [0.9, 1], y_1 > 1.5 \text{ kg/cm}^2, y_2 > 9\% \\ 0 (\text{劣质}), & \text{else} \end{cases} \quad (28)$$

3.5 结论

在本节中，我们针对问题 1 提出了两种定量模型：基于数据的线性回归和神经网络回归模型以及基于指标的层次分析法评价模型。

在回归模型中，我们拟合了机械指标与人工评分的关系，并利用多维正态分布生成的数据对两种指标进行了一致性分析，分析结果符合普遍认知，即机械指标较容易满足，而人工评分要求更高。

此外，基于回归模型我们提出了一种简单的定量分析模型，基于层次分析法我们提出了一种综合考量各指标的定量评价模型。与专家经验相比我们提出的模

型更为严格。

在基于回归模型的定量分析中，我们的模型与机械指标评判的重合率为 65.02%，而其中机械指标优质模型评价劣质的比例为 34.16%，证明回归模型对黄桃的优劣评价更为严格。

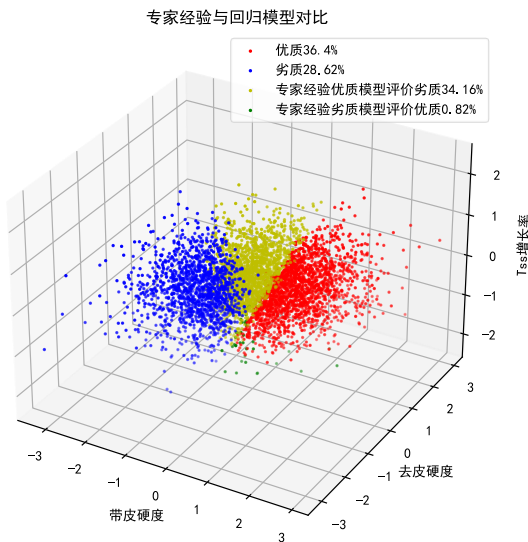


图 33 专家经验机械指标与回归模型对比

在基于层次分析法的定量分析中，我们的模型与机械指标评判的重合率为 55.73%。我们模型评价为优质的黄桃在机械指标都评价为优质，而机械指标评价为优质的黄桃在我们的模型中评价为劣质的比例为 44.26%。

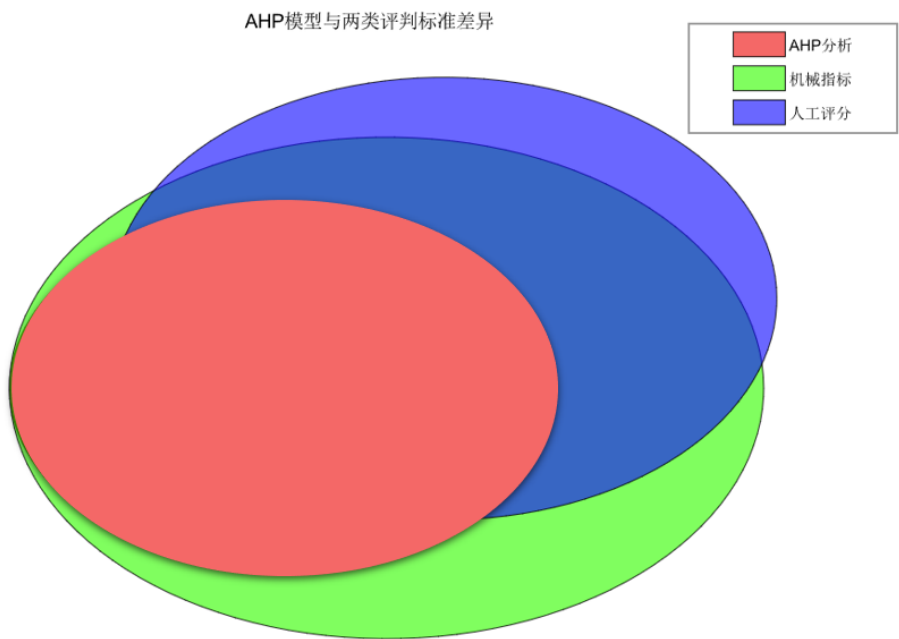


图 34 AHP 模型与两类评判标准差异

4 模型建立与求解

4.1 问题分析

在 2020 年度的评测数据中只提供了机械指标数据（包括带皮硬度、去皮硬度、TSS 含量和色差），没有提供人工评分。问题 2 中旨在利用 2020 年给出的机械指标数据，判断题中给出的 9 个样本的人工评分等级。自然地，我们会想到两种思路，其一是利用第三节中提到的回归模型，利用这些样本的机械指标预测出人工评分，再利用人工评分标准，即色泽、质地、味道均分大于 80 分给出对着 9 个样本的人工评分等级预测结果；其二是建立一个分类模型，对机械指标进行二分类，即分类为优质与非优质。

然而，我们还需要考虑一个问题，就是所给数据较少，训练出来的模型有可能出现过拟合或者欠拟合等情况，造成预测效果不佳。因此，我们提出了一种多模型联合预测机制，分别通过线性回归模型、SVM 分类模型、神经网络回归模型对数据进行拟合预测，再通过加权投票机制得到最终的预测结果。

4.2 基于机器学习的联合预测机制

我们提出的联合预测机制原理是利用不同预测模型通过机械指标对人工评分等级的结果进行加权投票，得到最终的投票结果。

我们使用的机械指标共有 9 项，分别是带皮硬度、去皮硬度、TSS 增长率和色差指标果皮 L、果皮 a、果皮 b、果肉 L、果肉 a、果肉 b。

设对第 i 个模型的加权值为 W_i ，对于加权系数 W_i 有

$$\sum_{i=1}^n W_i = 1, W_i \in [0,1] \quad (29)$$

设第 i 个模型的预测结果为 $Pred_i$ ，由于是二分类问题，因此 $Pred_i$ 的取值为 0, 1。设对各个模型的预测加权结果为 S ，而最终的预测结果为 $Pred$ ，有

$$S = \sum_{i=1}^n W_i Pred_i \quad W_i \in [0,1], Pred_i = 0,1 \quad (30)$$

$$Pred = u(S - 0.5) \quad (31)$$

其中 $u(\cdot)$ 表示阶跃函数，有

$$u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (32)$$

最终，我们的加权联合预测机制可由下图表示：

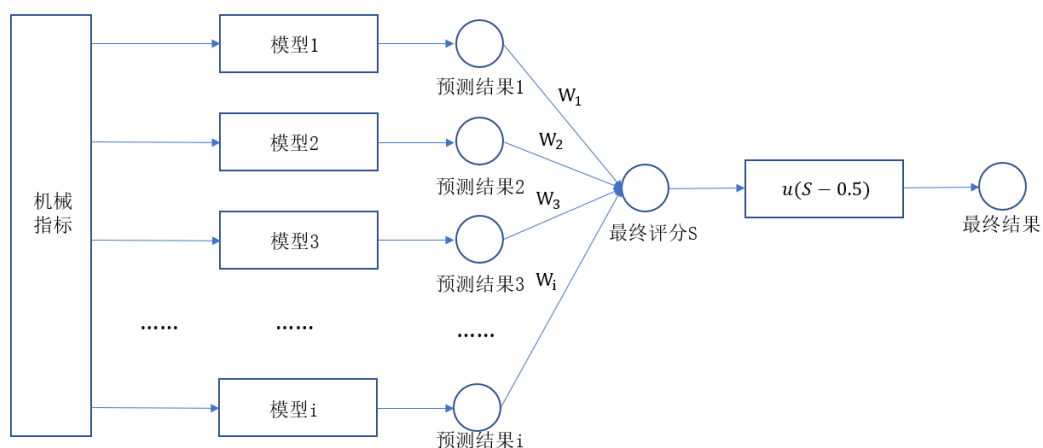


图 35 加权联合预测机制图

在本题中，我们采用的加权联合预测机制一共包括三种模型，分别是线性回归模型、SVM 分类模型、神经网络回归模型。

其中线性回归模型和神经网络回归模型在第三节中均一提到，在此不加赘述，这两个模型的可用下图表示：

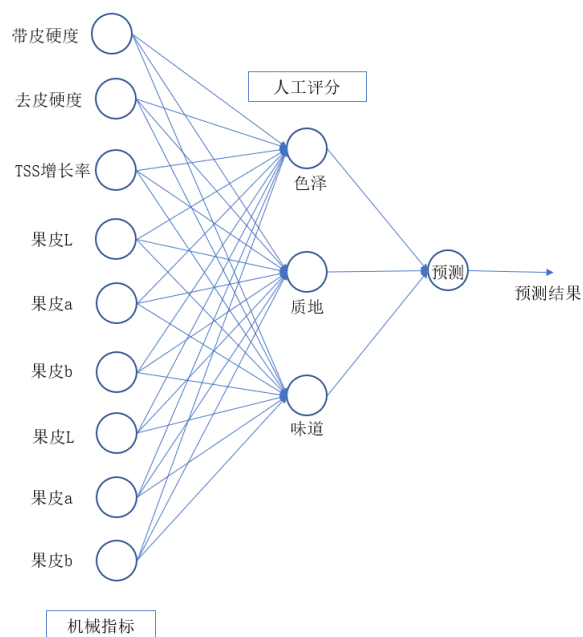


图 36 线性回归预测模型

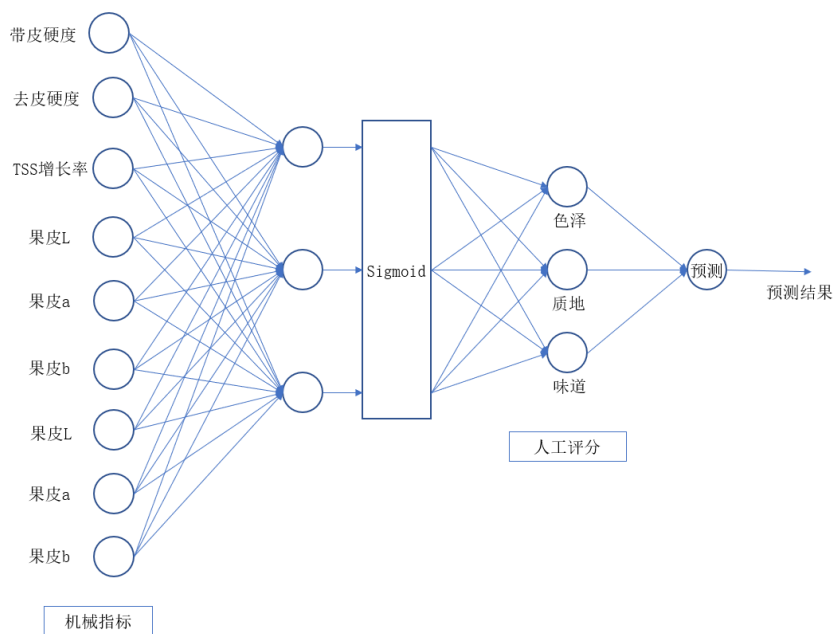


图 37 神经网络回归预测模型

除了线性回归模型和神经网络模型外，我们还引入了支持向量机（SVM）预测模型。SVM 是一种常见的二分类模型，其思想是找到一条分割曲线，使得类的数据点到该直线最近距离最大化，这些离直线最近的点就被称为支持向量。基于这个想法的目的在于使得这两个类离分类边界最远，若数据有微小的偏移也不会有太大影响。如下图，有“+”，“-”两类的数据点，SVM 的目的是希望找到如下中间的直线，使得间隔最大化。

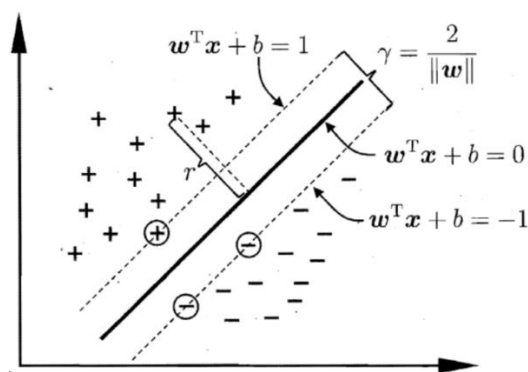


图 38 SVM 示意图

SVM 二分类模型具有适用于小样本数据、非线性映射的特点。引入 SVM 到我们的联合预测机制中，可与线性回归模型和神经网络模型互为补充。最终，我们的联合预测机制可如下图所示：

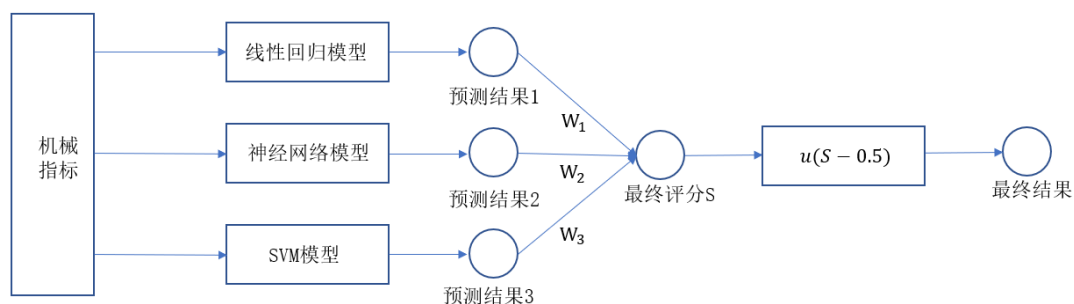


图 39 SVM 二分类预测模型

4.3 问题求解

问题中给出的 9 个样本及其机械指标如下表所示：

表 11 样本机械指标

样本编号	带皮硬度	去皮硬度	TSS 增长率	果皮 L	果皮 a	果皮 b	果肉 L	果肉 a	果肉 b
样本 A	7.41	5.05	0.18452047	83.7670	-7.2914	50.3956	83.2061	-2.658	59.8326
样本 B	1.45	1.24	0.15759955	78.6311	-0.4721	48.5373	86.3612	-0.9290	48.5268
样本 C	1.98	1.4	0.25630959	87.5126	-1.8342	54.993	88.4276	-0.1436	45.694
样本 D	3.83	2.15	0.10977080	81.1611	-3.1820	54.6889	79.1149	0.76403	61.9123
样本 E	1.49	1.05	0.03739445	83.0917	1.01525	52.4749	71.3038	3.06910	45.7971
样本 F	0.84	0.18	0.0293526	86.7303	1.95494	50.7718	72.9754	5.60196	49.1953
样本 G	2.79	1.74	0.1637931	87.1423	-0.1623	55.9013	65.1731	2.71543	45.4864
样本 H	1.73	0.33	0.1551724	88.5534	-2.7478	50.2095	81.7892	4.69566	52.7249
样本 I	0.98	0.65	0.2327586	85.0081	-0.0714	56.9654	75.6673	3.58742	49.7542

由表中我们可以看出，模型的输入为 9 维向量，分别代表带皮硬度、去皮硬度、TSS 增长率和色差指标果皮 L、果皮 a、果皮 b、果肉 L、果肉 a、果肉 b。

我们利用 2019 年的数据对线性回归模型、神经网络模型和 SVM 模型分别进行训练，得到拟合率分别为 0.625、0.75、0.77。

预测结果如下表所示：

表 12 预测结果表 (1)

模型 样本	线性回归预测	神经网络预测	SVM 预测
样本 A	1	1	1
样本 B	0	1	0
样本 C	1	0	0
样本 D	0	1	1
样本 E	0	0	0
样本 F	0	0	1
样本 G	0	0	0
样本 H	1	0	1
样本 I	0	0	0

根据拟合率，我们分别计算线性回归模型、神经网络模型和 SVM 模型的加权系数为

$$W_1 = \frac{0.625}{0.625 + 0.75 + 0.77} = 0.29 \quad (33)$$

$$W_2 = \frac{0.75}{0.625 + 0.75 + 0.77} = 0.35 \quad (34)$$

$$W_3 = \frac{0.77}{0.625 + 0.75 + 0.77} = 0.36 \quad (35)$$

通过加权投票机制，我们最终可以得到各个样本的预测结果和结果的置信度如下表所示：

表 13 预测结果图 (2)

样本	结果	置信度
样本 A	1	100%
样本 B	0	65%
样本 C	0	71%
样本 D	1	71%
样本 E	0	100%
样本 F	0	64%
样本 G	0	100%
样本 H	1	65%
样本 I	0	100%

综上所述，最终我们对 9 个样本的预测结果为样本 A、样本 D、样本 H 为优质黄桃。

4.4 结论

在本节中，我们采用基于投票的联合预测机制对问题 2 进行了求解，对题目中给出的 9 个样本，我们的预测结果是样本 A、样本 D、样本 H 为优质黄桃，置信度分别为 100%、71%、65%。

我们将结果与上表中的样本机械指标对比可以发现：若按照题目中给定的由机械指标判断优质与非优质黄桃的标准，入选优质的黄桃样本应该是样本 A、样本 D、样本 G，与我们的结果有 2/3 的吻合度。但如果我们将色差指标纳入考虑范围，则会发现样本 G 的果肉 L 指标过低，不符合第三节中层次分析法提到的优质黄桃应有的 L 指标范围。而我们所预测到的样本 H，虽然去皮硬度较低，然而其他指标（包括带皮硬度、TSS 增长率、色差指标）都符合要求。从 2018 年和 2019 年的数据中我们也可以发现，存在机械指标判断为劣而人工指标判断为优的情况，因而我们推测样本 H 有可能是人工评级为优而机械评级为非优的样本。

5 模型建立与求解

5.1 问题分析

建立黄桃品质随贮藏环境和贮藏时间变化的预测模型,对选择合适的贮藏方式和时间具有重要意义,一般通过两类方法来构建:基于原理的微分方程预测和基于数据的函数拟合和预测。黄桃的品质变化其根本原因是内部化学反应的发生,因此构建化学动力学模型来进行品质变化的拟合是合乎逻辑的。另一方面,也可以通过数据直接拟合黄桃指标随时间变化趋势达到预测的目的。常用的拟合方法有基于多项式、幂函数、傅里叶系数、样条差值方法的拟合,但是这些拟合方式往往只能对单变量进行拟合,而考虑不到各个变量之间相互关系,因此我们也采用长短记忆网络来进行多指标变化的拟合和预测。但是基于循环神经网络的方法是需要大量数据的,而每一批黄桃的测试次数非常有限,因此对数据首先进行插值也是很有必要的,为了简易判断桃子的优质状态,以及利用尽可能多的数据,我们取三年的数据,只收集带皮硬度、去皮硬度和 TSS 下降率,一共有 18 批数据,本节我们以每一批黄桃为研究对象。

5.2 基于化学反应动力模型的预测

5.2.1 反应方程建立

尝试采用化学反应动力学模型来拟合和预测黄桃贮藏过程中的品质变化。大量试验表明,大多数果蔬的指标在贮藏过程中遵循零级或一级反应模型,其动力学方程如下所示:

$$\text{零级反应: } I_0 - I = kt \quad (36)$$

$$\text{一级反应: } \ln(I_0) - \ln(I) = kt \quad (37)$$

其中 t 为贮藏时间, I_0 为黄桃贮藏初始某项机械指标, I 为贮藏 t 天后的某项机械指标, k 为反应的速率常数。根据黄桃贮藏中随时间变化的指标数据,可以确定反应级数。

Arrhenius 提出了反映温度 T 与反应速率常数 k 的 Arrhenius 公式, 如下:

$$k = Ae^{-E_a/RT} \quad (38)$$

对其变形得:

$$\ln k = -\frac{E_a}{RT} + \ln A \quad (39)$$

其中 A 为指前因子, E_a 为实验活化能, R 为气体常数 ($8.314\text{J/mol}\cdot\text{K}$), T 为热力学温度。若确定反应级数方程和反应速率常数, 以 $\ln k$ 对 $1/T$ 作图, 通过其斜率和横轴截距可以计算出指前因子 A 和活化能 E_a 的数值, 从而构建预测模型。

5.2.2 化学反应方程求解

以 2018 年 6 月 11 日冷柜 (8°C) 贮藏抽取的黄桃为例, 得到其机械指标变

化趋势及相关参数如下：

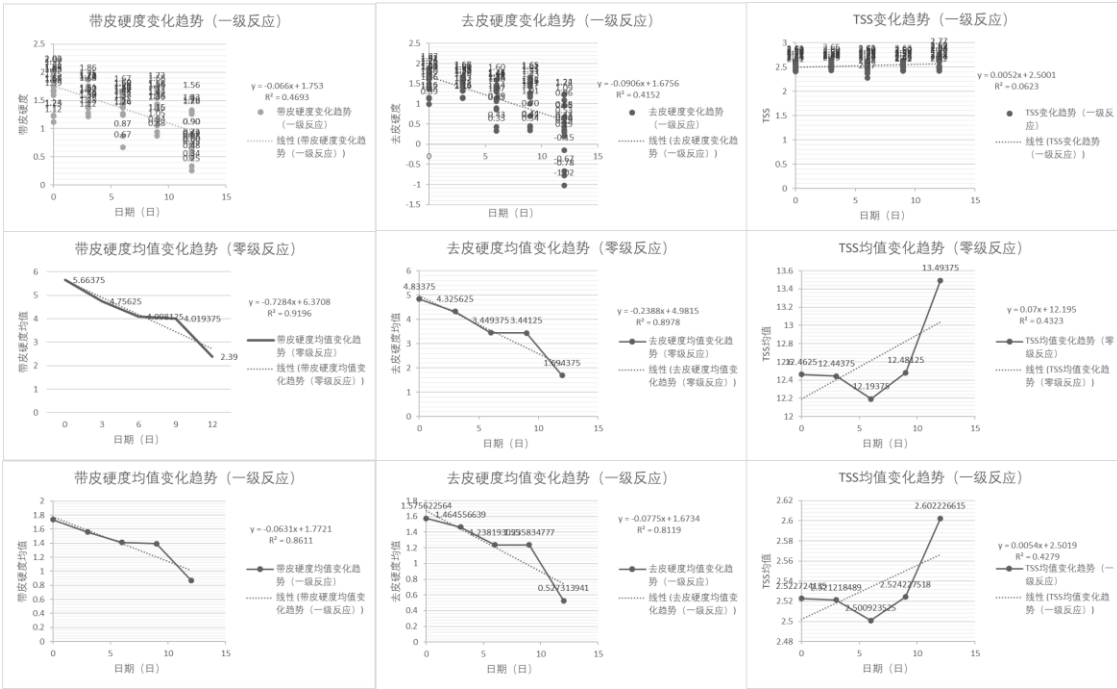


图 40 2018 年 6 月 11 日批次冷柜 (8℃) 黄桃硬度及 TSS 变化趋势图
处理 2018 年所有批次黄桃样本的机械指标得到下表：

表 14 2018 年黄桃样品反应速率常数及相关系数

贮藏条件	批次及指标	零级反应		一级反应	
		反应速率常数 k	相关系数 R^2	反应速率常数 k	相关系数 R^2
冷柜 (8℃), T=281.15K	6.11 批次带皮硬度	0.7284	0.9196	0.0631	0.8611
	6.11 批次去皮硬度	0.2388	0.8978	-0.0775	0.8119
	6.11 批次 TSS	-0.0700	0.4323	0.0054	0.4279
	6.14 批次带皮硬度	0.8641	0.8739	0.0582	0.8946
	6.14 批次去皮硬度	0.2090	0.8609	0.0544	0.8580
	6.14 批次 TSS	-0.1212	0.7529	-0.0103	0.7571
	6.16 批次带皮硬度	0.5448	0.7197	0.0646	0.7642
	6.16 批次去皮硬度	0.1439	0.6346	0.0791	0.6860
	6.16 批次 TSS	-0.1708	0.8918	-0.0146	0.8833
冷库 (1℃),	6.11 批次带皮硬度	-0.4766	0.7205	-0.0167	0.727

T=274.15K	6.11 批次去皮硬度	-0.0726	0.9683	-0.0161	0.9582
	6.11 批次 TSS	-0.0028	0.0039	-0.0002	0.0040
	6.14 批次带皮硬度	-0.2613	0.3447	-0.0080	0.3605
	6.14 批次去皮硬度	-0.0346	0.4639	-0.0076	0.4783
	6.14 批次 TSS	0.0029	0.0025	0.0003	0.0026
	6.16 批次带皮硬度	0.0884	0.0450	0.0035	0.0543
	6.16 批次去皮硬度	0.0128	0.0299	0.0039	0.0390
	6.16 批次 TSS	-0.0151	0.0459	-0.0014	0.0524

从表中可以看出，黄桃在贮藏环境为冷柜（8℃）的情况下，其硬度和 TSS 比较符合零级反应方程，但是在冷库贮藏环境下除去 6 月 11 日批次的其余两个批次零级反应和一级反应相关系数不显著，说明拟合效果很差。在贮藏环境为冷库时，甚至出现反应速率常数 k 为负值的情况，这说明指标总体呈现随贮藏时间增长而增加的趋势，说明贮藏温度对于黄桃品质的影响很大。

对于冷柜（8℃），我们根据零级反应方程和 2018 年的各批次反应速率常数可以简单构建线性预测模型 $I = I_0 - kt$ 。

带皮硬度的线性预测模型为： $I_1 = 5.7700 - 0.7138t$

去皮硬度的线性预测模型为： $I_2 = 4.4946 - 0.1972t$

TSS 线性预测模型为： $I_3 = 11.3871 + 0.1207t$

我们从 2019 年贮藏方式为冷柜的每个批次中任选一个数据来检验预测的准确性。结果如下表：

表 15 预测模型误差率

数据序号	实际带皮硬度	预测带皮硬度	误差率	实际去皮硬度	预测去皮硬度	误差率	实际 TSS	预测 TSS	误差率
1	4.7850	4.3424	9.24%	4.5955	4.1002	10.78%	12.65	11.6285	8.07%
2	3.1050	2.9148	6.13%	3.4850	3.7058	6.33%	12.85	11.8699	7.63%
3	1.5500	1.4872	4.05%	1.0350	3.3114	219.94%	10.05	12.1113	20.51%

可以看出大部分情况预测误差率在 10%以内，预测效果较为可以接受。但是也存在少量数据预测结果差的情况，如上表中误差率甚至有 219.94%。对出现这种情况进行分析：

1) 可能存在某个样品在运送去贮藏的路程中或者在贮藏过程中损坏，导致样品指标数据与其余大部分样品指标数据差异明显。

2) 因为构建这个线性预测模型时初始指标数据 I_0 是取 2018 年贮藏于冷柜三

个批次的初始黄桃样品平均值，之前分析得出温度对于黄桃的品质有着很大影响，而三个批次黄桃入柜贮藏的初始日期不一样，而该线性预测模型忽略了贮藏前常温环境对于黄桃的影响以及黄桃未采摘下时桃树对于黄桃品质的影响，因此可能导致误差率较大。

再考虑动力学方程

$$\text{零级反应: } I_0 - I = kt \tag{40}$$

$$\text{一级反应: } \ln(I_0) - \ln(I) = kt \tag{41}$$

$$k = Ae^{-E_a/RT} \tag{42}$$

通过方程中可以看出， k 是关于温度 T 的指数函数， k 值的正负取决于指前因子 A 的正负，而指前因子是一个定值，因此 k 值的正负是确定的。从 2018 年样品黄桃反应系数表中可以得知在冷柜贮藏条件下指标的 k 值多为正值，在冷库贮藏条件下指标的 k 值多为负值，并且相关性很差，因此无法应用动力学方程来拟合和预测 $1^{\circ}\text{C} \sim 8^{\circ}\text{C}$ 贮藏温度时黄桃品质变化。

5.2.3 数据分析初步预测采摘时间

从线性预测模型中可以看出贮藏前的常温环境和采摘时间对贮藏时黄桃的初始品质影响很大。为了把握一个较好的采摘时机，对贮藏各批次黄桃的初始品质按采摘时间进行统计。

黄桃在运送去贮藏的过程中可能有损伤或者机械测评时有差错，为了减小这种异常数据带来的影响，我们采用中位数来观察各批次初始样品指标：

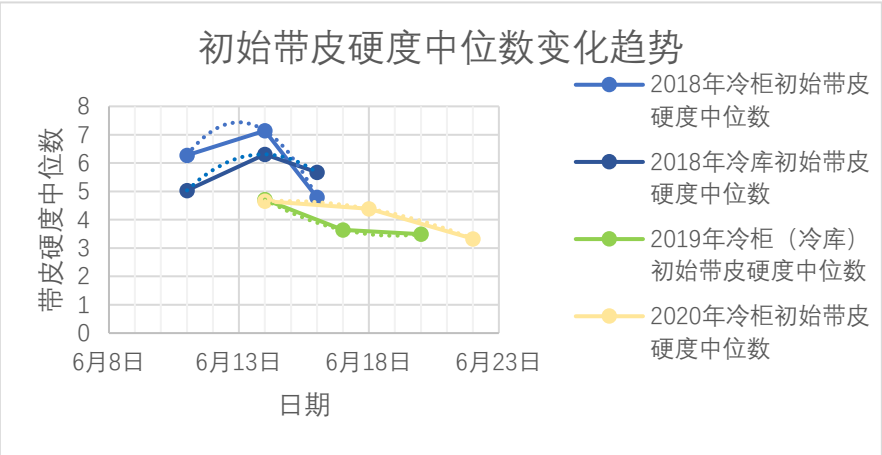


图 41 初始带皮硬度中位数变化趋势图

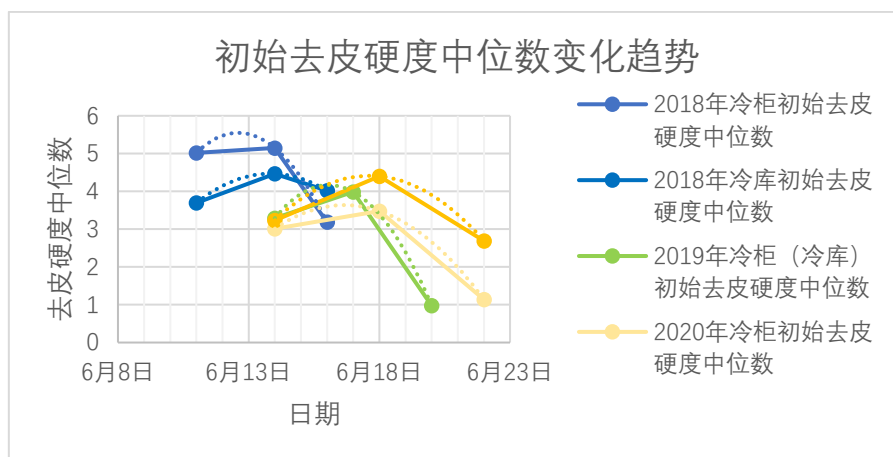


图 42 初始去皮硬度中位数变化趋势图

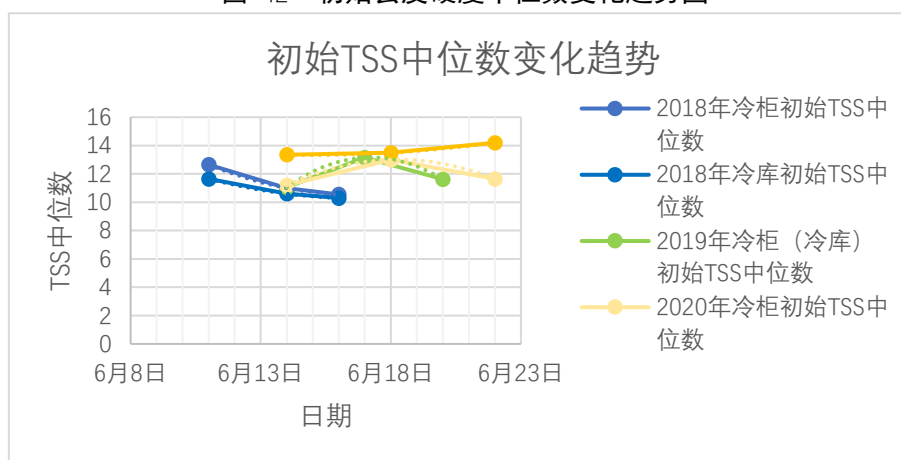


图 43 初始 TSS 中位数变化趋势图

通过用多项式拟合变化曲线，我们可以归纳每年各项指标的峰值出现的大致时间。由于每年成熟期不一定一致，我们假设每年第一批次采摘当天是该年度黄桃刚刚进入成熟期的第一天，以进入成熟期的天数来表示峰值时间。

表 16 各批次指标峰值出现时间

年份	带皮硬度峰值时间/日	去皮硬度峰值时间/日	TSS 峰值时间/日
2018	2~3	3~4	1~2
2019	1	3~4	3~4
2020	1	3~4	9

黄桃进入成熟期后可挂树 10 天左右，从以上图中可以看出在常温环境下黄桃挂树时，刚进入成熟期 1-2 天，带皮硬度最高；在进入成熟期 3-4 天左右，带皮硬度显著降低，而此时去皮硬度处于较高水平；在进入成熟期 7-8 天左右，去皮硬度明显降低。而 TSS 从图表中无法得到变化规律，TSS 指的是可溶性糖类，主要是蔗糖、葡萄糖和果糖。而挂树果实中可溶性糖类受当月降雨量、温度、日照和土质的影响较大，黄桃的 TSS 会产生波动^[7]，但根据对本题数据的分析，相较于硬度的变化量，TSS 的变化量很少能够影响到对于黄桃品质的评判。

我们再统计每一批次根据机械指标测量首次出现优质率小于 50%时的贮藏天数。如果优质率没有小于 50%，就记录最长贮藏天数。

表 17 优质率与贮藏天数

批次	样品机械指标优质率	贮藏天数	对应样品人工评分 优质率
2018 年 6 月 11 冷柜	50%	12	66.7%
2018 年 6 月 11 冷库	93.75%	20	91.7%
2018 年 6 月 14 冷柜	93.75%	12	27.3%
2018 年 6 月 14 冷库	100%	20	12.5%
2018 年 6 月 16 冷柜	37.5%	3	87.5%
2018 年 6 月 16 冷库	93.75%	20	46.2%
2019 年 6 月 14 冷柜	40%	8	87.5%
2019 年 6 月 14 冷库	80%	20	60%
2019 年 6 月 17 冷柜	40%	4	100%
2019 年 6 月 17 冷库	0%	18	78.8%
2019 年 6 月 20 冷柜	20% (40%)	0 (2)	87.5% (100%)
2019 年 6 月 20 冷库	20% (40%)	0 (10)	87.5% (44.4%)
2020 年 6 月 14 冷柜	40%	8	/
2020 年 6 月 14 冷库	60%	25	/
2020 年 6 月 18 冷柜	40%	4	/
2020 年 6 月 18 冷库	90%	25	/
2020 年 6 月 22 冷柜	45% (40%)	0 (2)	/
2020 年 6 月 22 冷库	84.6%	25	/

通过对上表数据分析，在冷库（1℃）贮藏条件下，保质期较长，即在保证机械指标优质率的前提下，能够保存更长时间，但是人工评分的优质率却相较较差，意味着风味确实有所下降。而在冷柜（8℃）贮藏条件下，黄桃的保质期较短，但人工评分的优质率却相较较高，表示冷柜贮藏短时间内能够较好地保存风味。

通过以上分析，可以对黄桃的采摘时间和贮藏方式进行初步预测。建议黄桃的采摘时间在进入成熟期后的 3-4 天。如果想在贮藏一段时间后售卖时保持较好的风味，可以在冷柜（8℃）的环境中贮藏 8-14 天；如果想延长贮藏时间，可以在冷库（1℃）的环境中贮藏 18-20 天左右。

5.2 基于数据的预测

5.2.1 传统曲线拟合

传统的曲线拟合方式有很多，例如基于多项式的拟合、基于指数函数的拟合，基于傅里叶序列的拟合还有基于平滑样条的拟合，我们对黄桃指标随时间趋势用以上方法进行了拟合，对比分析结果如下：

（1）基于多项式的拟合

使用多项式的拟合方程如下：

$$y = \sum_{i=1}^{n+1} p_i x^{n+1-i} \quad (43)$$

其中 $n+1$ 为多项式模型的项数， n 是多项式模型的阶数，拟合情况如下：

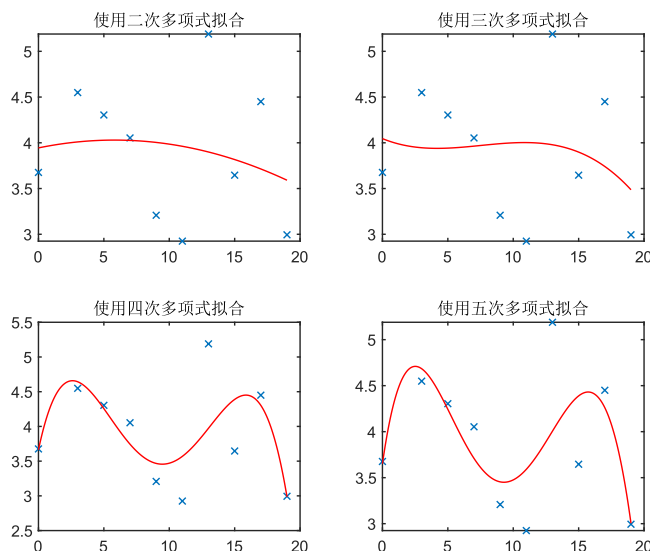


图 44 多项式拟合

可以看到不同阶多项式拟合对该数据具有合理的灵活性，并且它们是线性的，所以拟合过程很简单，但是高度拟合会变得不稳定，多项式可以在数据范围内提供良好的拟合，但在该范围之外可能会发散很大，所以不适宜外推。

(2) 指数函数拟合

指数函数模型如下：

$$y = ae^{bx}$$

$$y = ae^{bx} + ce^{dx} \quad (44)$$

使用指数进行拟合一般采用一项或者两项，拟合结果如下：

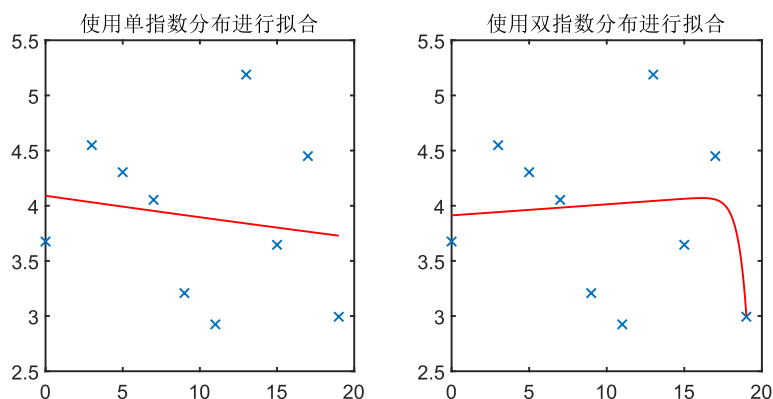


图 45 指数函数拟合

可以看到两项以内的指数函数对序列欠拟合，然而使用过多项进行外推，拟合曲线会很大无法收敛，因此此方案不能使用。

(3) 傅里叶序列拟合

傅立叶级数是描述周期信号的正弦和余弦函数的总和。它以三角函数形式或指数形式表示，模型如下：

$$y = a_0 + \sum_{i=1}^n a_i \cos(iwx) + b_i \sin(iwx) \quad (45)$$

其中 n 是傅里叶序列阶数， a 与 b 分别是正弦项和余弦项的系数，拟合情况如下：

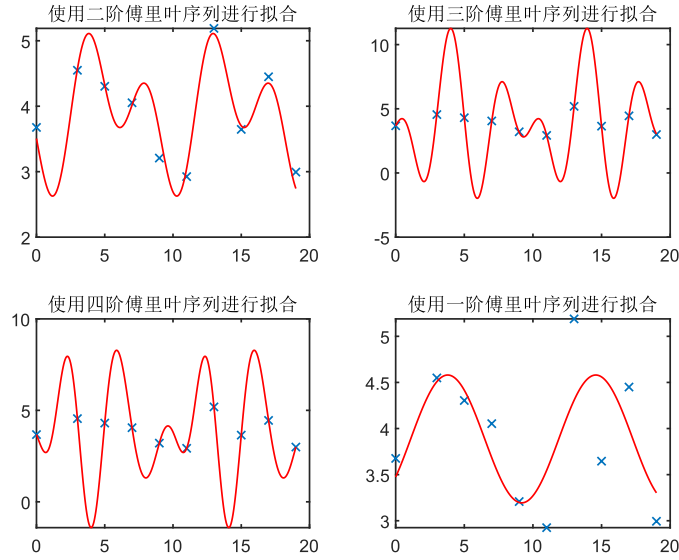


图 46 傅里叶序列拟合

可以看到傅里叶序列拟合基本上在样本点上没有误差，但是由于序列本身性质，拟合曲线具有一定周期性，这与机械指标变化过程有所违背，因此不适合采用。

(4) 平滑样条插值

样条是以分段多项式函数得来的，三次平滑样条拟合需要最小化一下目标值：

$$loss = p \sum_{j=1}^n w_i |y_i - f(x_j)|^2 + (1-p) \int \lambda(t) |D^2 f(t)|^2 dt \quad (46)$$

其中左边一项是误差量度，右边一项是粗糙量度，目的是使样值点误差最小，拟合函数越平滑，而 p 是两者的权重指标，拟合结果如下：

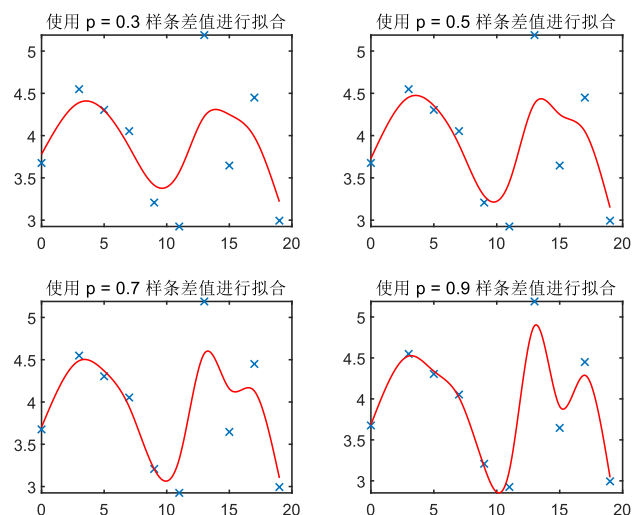


图 47 平滑样条插值

可以看到随着 p 的增加，函数在样值点的误差减小，但是曲线的平滑度也在下降， $p=0.7$ 的样条函数能够很好兼顾准确性和平滑性，因此我们采用该插值方法对指标变化情况进行拟合和外推。

(5) 采用平滑样条对机械指标插值结果

对冷库和冷柜三年各九批次黄桃三个指标进行插值结果如下：

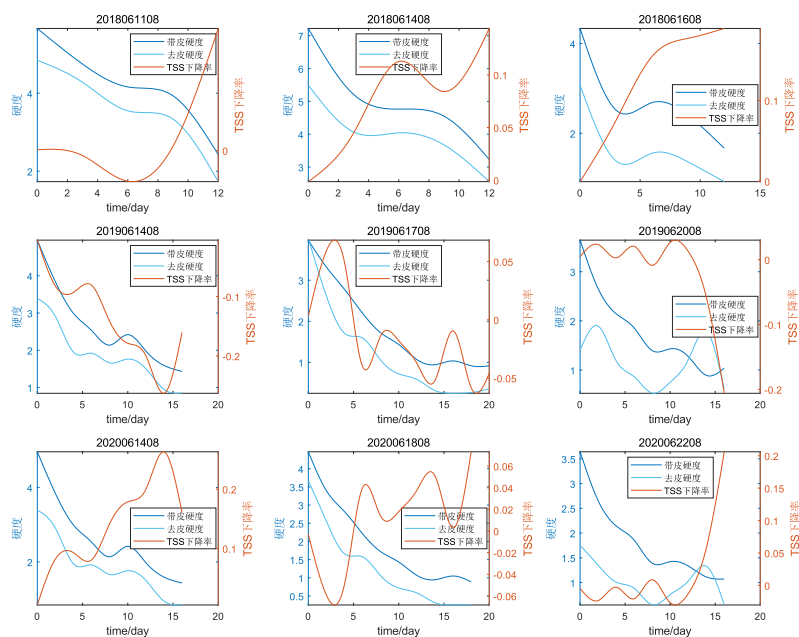


图 48 冷柜 (8°C) 机械指标随时间变化情况

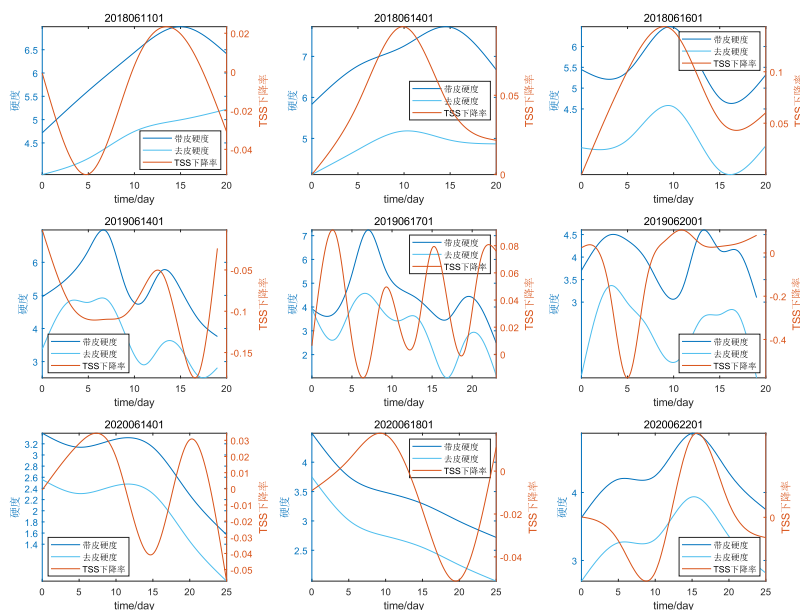


图 49 冷库（1℃）机械指标随时间变化情况

5.2.2 长短时记忆网络对多指标同时拟合

采用平滑样条可以得到比较好的效果，但是只能对一个指标进行预测，无法兼顾各个指标之间的关系。因为机械指标前后具有很强的关联性，不同指标相互之间可能会有作用，因此使用具有记忆和使用历史信息的能力的循环卷积网络进行拟合可能会获得比较好的效果，具体使用的是长短记忆网络（LSTMnet），其中 LSTM 块结构如下：

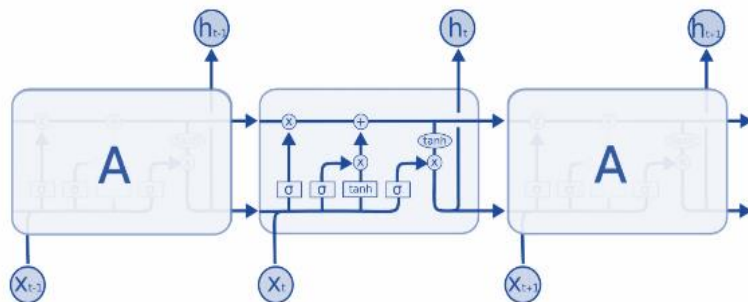


图 50 LSTM 块结构

LSTM 适合于处理和预测时间序列。最左边函数为区块的 input，右边三个会经过 gate 决定 input 是否能传入区块，左边第二个为 input gate，如果这里产出近似于零，将这里的值挡住，不会进到下一层。左边第三个是 forget gate，当这产生值近似于零，将把区块里记住的值忘掉。第四个也就是最右边的 input 为 output gate，它可以决定在区块记忆中的 input 是否能输出。整个网络的结构为：



LSTM 网络训练参数与训练过程为:

表 18 LSTM 网络训练参数表

超参名	值
最大代数	200
梯度阈值（防止梯度爆炸）	1
初始学习率	0.005
学习率规划	分段
优化方式	ADAM
学习率初段周期	150

超参名	值
最大代数	200
梯度阈值（防止梯度爆炸）	1
初始学习率	0.005
学习率规划	分段
优化方式	ADAM
学习率初段周期	150

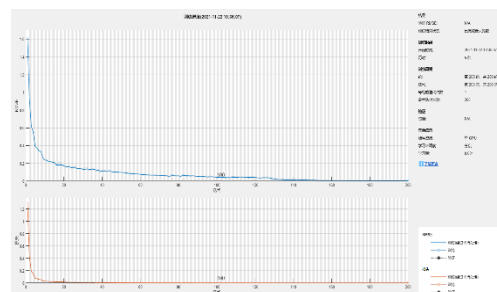


图 51 LSTM 网络训练过程

实验结果:

取 2018 年 6 月 16 日采摘的杨桃, 观测其三个机械指标随时间变化情况, 并使用训练好的 LSTM 输入前 80 数据, 对后 20 数据进行预测, 结果如下:

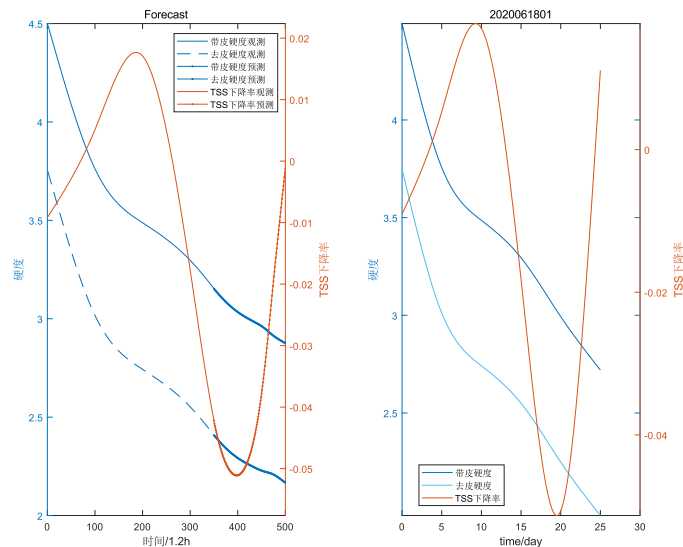


图 52 LSTM 预测结果图 (1)

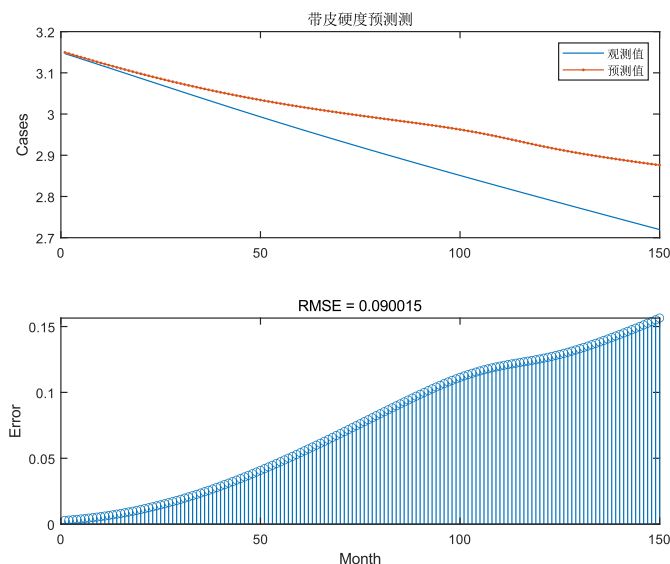


图 53 LSTM 预测结果图 (2)

可以看到 LSTM 网络可以较好的对三个指标同时进行预测，不过预测误差，随着时间有积累的趋势，并且对动态范围较小的指标如 TSS 下降率的预测相对误差较大。这说明网络的训练是过拟合的，根本原因是数据量较小，采用插值的方法增强数据并不能降低网络的过拟合程度。

5.2.3 基于数据预测模型的结论

(1) 合适的储藏方式与时间

本次实验采用机械指标，因此评判标准均以果实硬度 $>1.5\text{kg}/\text{cm}^2$, TSS 含量下降率 $<20\%$ 为优质。通过对 18 批黄桃的指标随时间变化的预测来看，不仅存储环境对果实指标变化很大，初始值也有较大影响，为了聚焦存储环境的影响，我们把冷库和冷柜各九批数据预测结果平均作为该存储环境下指标的变化情况：

表 19 指标合格天数预测

指标合格 时间/天 存储环境	TSS 下降率	带皮硬度	带皮硬度	推荐存储
冷库（1℃）	29	31	25	25
冷柜（8℃）	22	18	14	14

可以看到在冷库中三项指标到达临界所需要时间均比冷柜中长，因此冷库存储保质期长，口味可以从 TSS 下降率谷值来判定，从 5.2.1 中 5 小结可以看到，冷柜中 TSS 下降率最小值普遍低于冷库对应值，TSS 下降率小意味着 TSS 增长峰值高，这说明在冷柜中，黄桃的代谢较高，呼吸作用强度高，糖分产量高。

（2）需要进行多少次有损测试

我们 2018 年 6 月 22 日采摘并置入冷柜的那批黄桃为例为例，假设每 2 天进行一次测试，并通过测试结果更新网络，得到黄桃带皮硬度指标变化情况，再与实际值对比可以得到预测的误差均方差，显然预测次数越多，预测均方差就越小，我们可以得到预测次数和预测均方误差的关系图如下：

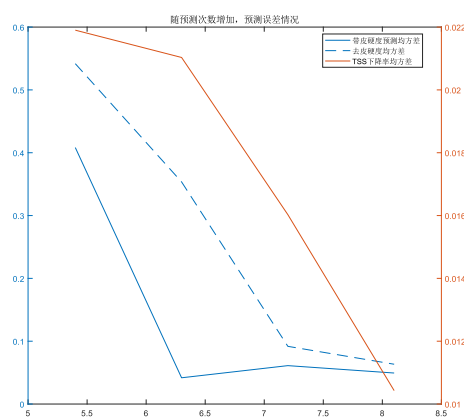


图 54 预测次数和预测方差关系图

通过该图可以看到误差随着预测次数增加的降低，在本情况下，预测 6 次较为合理。

6 模型评估

6.1 模型的优点

问题一

我们的模型在考虑了各个年份数据存在不少异同之处的条件下，充分利用了题目中提供的数据。在解决两种指标的样本不对应问题时，采用了批平均操作。

在对两种指标进行一致性分析和建立定量模型时，使用了线性回归分析和非线性神经网络回归，充分考虑了数据间存在的线性性与非线性性。

基于层次分析法构建的评价模型能够容纳更多的评价指标，将机械指标中的

色差和人工评分中的芳香也加入到模型中,综合机械指标和人工评分的分数对黄桃品质作出更全面地评价。

问题二

在解决问题二时,本文创造性地提出了加权联合预测机制,利用线性分析预测、神经网络预测和 SVM 预测对 2019 年的数据进行拟合并对 2020 年的给定样本进行人工评级预测。利用适用于线性、非线性、小样本的三种模型对样本进行综合评判,缓解了由于数据量少带来的过拟合和欠拟合问题。

问题三

基于化学反应动力方程式的预测模型构建较为简单,考虑了果实贮藏时环境温度 and 贮藏天数等因素。

通过散点图、直方图估计,平行坐标轴的构建清晰可视化了黄桃各指标关系,并用 WPCA 的方法进一步深化了对指标冗余度以及指标关联性的研究,结果清晰明了。

通过三次平滑样条对桃子品质变化曲线进行拟合,再用 LSTM 网络进行了多个指标联合预测,考虑了多个指标关系。

6.2 模型的缺点

问题一

基于层次分析法构建的评价模型在对色差进行分类分析时使用目视比色法,可能存在判断误差导致色样集分类不准确。在用层次分析法比较同一层次不同因子的影响权重时调查对象数量不足,构建的判断矩阵可能不具有普遍代表性。

问题二

由于数据量较少,模型对于所给数据的拟合程度不够高,如果增加训练的数据量或许能够提高效果。

问题三

由于在本题中黄桃所处环境有常温、冷柜(8℃)和冷库(1℃),进入贮藏环境后其品质变化趋势与初始采摘时间相关,而经过分析基于化学反应动力方程式的预测模型模型无法适应这种相关变化,不能较好拟合和预测冷库贮藏条件下黄桃品质的变化趋势。

散点图、WPCA 等数据观测分析方法,只能做到对数据的线性关系挖掘,未能充分发掘潜在关系。

样条插值可以对数据进行插值但是外推时具有很大风险,三年一共只有 18 批桃子,数据量太小, LSTM 网络容易过拟合,并且, LSTM 是隐式特征提取,对问题概念分析、以及可解释性较差。储藏时间以及测试次数只能通过实验的方式取平均来估计。并且实验发现对单样本进行训练效果较好,但是对同一储存环境下所有样本进行训练得到的网络性能较差,这可能是因为即使是同一存储环境,黄桃性能变化曲线仍然差异很大。这可能与曲线初始值关系很大,然而在训练前已经将样本归一化无法体现。

参考文献

- [1]刘莉. 黄桃罐头感官品质评价研究[D].天津科技大学,2015.
- [2] 颜色标准 LAB 值对照表.<https://wenku.baidu.com/view/a6b30c4d9f3143323968011ca300a6c30d22f12c.html>.2021 年 11 月 20 日.
- [3] NY/T844-2017.中华人民共和国农业行业标准.绿色食品 温带水果.
- [4] NY/T586-2002.中华人民共和国农业行业标准.鲜桃.
- [5]DB31/T488-2019.上海市地方标准.地理标志产品.奉贤黄桃.
- [6]《锦绣黄桃质量等级标准》标准文本（征求意见稿）.<https://www.renrendoc.com/paper/99715841.html>.2021 年 11 月 20 日.
- [7]滕玉香, 刘芬, 谢孝东. 麻阳县黄桃产量和可溶性固形物含量与气象条件的相关性分析[J]. 现代农业科技, 2019.

附录

问题编号	代码名称	代码作用
问题 1	2018 线性回归拟合.py	拟合两种指标以及一致性分析
	2019 线性回归拟合.py	拟合两种指标以及一致性分析
	2018 神经网络拟合.py	拟合两种指标以及一致性分析
	2019 神经网络拟合.py	拟合两种指标以及一致性分析
	层次分析法.py	利用层次分析法进行定量分析
	w pca.mlx	WPCA 指标间关系分析
问题 2	2020 年线性回归预测.py	线性回归模型预测样本优劣
	2020 年神经网络预测.py	神经网络模型预测样本优劣
	2020 年 SVM 分类预测.py	SVM 分类模型预测样本优劣
问题 3	差值拟合.mlx	差值拟合工具包
	LSTM 预测.mlx	LSTM 预测
其他	readData3.py	数据读取工具包
	Utils.py	其他工具包
	Model.py	储存模型类
	可视化.mlx	可视化工具包
	venn.m	维恩图工具包

核心代码：

1. 线性回归分析（python）

```

class KB(nn.Module):
    def __init__(self, in_ch, out_ch):
        super(KB, self).__init__()
        self.k = nn.Parameter(torch.randn(in_ch, out_ch, dtype=torch.double))
        self.b = nn.Parameter(torch.randn(1, out_ch, dtype=torch.double))
    def forward(self, X):
        y = torch.matmul(X, self.k)
        y = torch.add(y, self.b)
        return y

model = KB()
lossFunc = nn.MSELoss()
optim = torch.optim.Adam(model.parameters(), lr=0.001)
X = torch.tensor(X)
Y = torch.tensor(Y)
for epoch_i in range(100000):
    optim.zero_grad()
    pred = model(X)
    loss = lossFunc(Y, pred)
    loss.backward()
    optim.step()

```

```
print("epoch: {0}, loss: {1}".format(epoch_i, loss.detach().numpy()))
```

2. 神经网络回归（python）

```
class FC(nn.Module):
    def __init__(self, in_ch, mid_ch, out_ch):
        super(FC, self).__init__()
        self.liner1 = nn.Linear(in_ch, mid_ch)
        self.relu = nn.ReLU()
        self.dropout = nn.Dropout()
        self.liner2 = nn.Linear(mid_ch, out_ch)
        self.sigmoid = nn.Sigmoid()
    def forward(self, x):
        x = self.liner1(x)
        x = self.sigmoid(x)
        #x = self.dropout(x)
        x = self.liner2(x)
        y = self.sigmoid(x)
        return y
model = FC(in_ch=3, mid_ch=3, out_ch=3)
lossFunc = nn.MSELoss()
optim = torch.optim.Adam(model.parameters(), lr=0.001)
X = torch.tensor(X, dtype=torch.float32)
Y = torch.tensor(Y, dtype=torch.float32)
for epoch_i in range(10000):
    optim.zero_grad()
    pred = model(X)
    loss = lossFunc(Y, pred)
    loss.backward()
    optim.step()
    print("epoch: {0}, loss: {1}".format(epoch_i, loss.detach().numpy()))
```

3. 层次分析法（python）

```
def WeiScore(MechanicalIndexArray, ManualScoreArray, std_sca, man_std_sca):
    preMechanicalIndexArray = MechanicalIndexArray
    MechanicalIndexArray = np.empty((preMechanicalIndexArray.shape[0],
preMechanicalIndexArray.shape[1] - 1))
    MechanicalIndexArray[:, :-1] = preMechanicalIndexArray[:, :-2]
    ones = np.ones(MechanicalIndexArray.shape[0])
    zeros = np.zeros(MechanicalIndexArray.shape[0])
    MechanicalIndexArray[:, -1] = np.where(
        (preMechanicalIndexArray[:, -2] >= 80) & (preMechanicalIndexArray[:, -1] >
50), ones, zeros)
```

```
MechanicalIndexArray[:, :3] =
```

```

std_sca.inverse_transform(MechanicalIndexArray[:, :3])
ManualScoreArray = man_std_sca.inverse_transform(ManualScoreArray)

S = ManualScoreArray[:, 0]*0.1495 + ManualScoreArray[:, 1]*0.2402 +
ManualScoreArray[:, 2]*0.2480 + ManualScoreArray[:, 3]*0.3623
M_Standard = np.array([1.5, 1.5, -0.2, 0.5])

M = np.ones(MechanicalIndexArray.shape[0])
neg_idx = set(np.where((MechanicalIndexArray - M_Standard) < 0)[0])
M[np.array(list(neg_idx))] = 0

F = M + S/100

label = np.zeros(F.shape[0])
label[np.where(F>1.75)[0]] = 1

label[np.where((F>=0.9)&(F<=1)&(MechanicalIndexArray[:,0]>1.5)&(MechanicalIndexArray[:,1]>1.5)&(MechanicalIndexArray[:,2]>0.09))][0]] = 1
return F, label
cube = createCube(np.concatenate([ManualScoreArray, MechanicalIndexArray], 1))
F, labelWei = WeiScore(cube[:, 4:], cube[:, :4], std_sca, man_std_sca)
man_pos, man_neg, labelMan = ManScore(cube[:, :4], man_std_sca)
mec_pos, mec_neg, labelMec = MecScore(cube[:, 4:], std_sca)

man_pos_set = set(np.where(labelMan == 1)[0])
mec_pos_set = set(np.where(labelMec == 1)[0])
wei_pos_set = set(np.where(labelWei == 1)[0])

man_neg_set = set(np.where(labelMan == 0)[0])
mec_neg_set = set(np.where(labelMec == 0)[0])
wei_neg_set = set(np.where(labelWei == 0)[0])

len(mec_pos_set&man_neg_set&wei_neg_set)

```

4. 问题 2 预测代码 (python)

```

X = np.array(X)
Y = np.array(Y)
Standard = (np.array([[1.5, 1.5, -0.2]])-std_sca.mean_[:3])/std_sca.scale_[:3]

Standard_man = man_std_sca.transform(np.array([[80, 80, 80]]))

neg_idx = set(np.where((X[:, 1:4]-Standard) < 0)[0])
pos_idx = set(np.arange(X.shape[0])) - neg_idx

```

```
neg_man_idx = set(np.where(np.mean(man_std_sca.inverse_transform(Y[:, 1:4])-
man_std_sca.inverse_transform(Standard_man), 1) < 0)[0])
pos_man_idx = set(np.arange(Y.shape[0])) - neg_man_idx
```

```
label = np.zeros((X.shape[0], 1))
label[list(pos_man_idx), :] = 1
```

```
model = FC(in_ch=9, mid_ch=3, out_ch=3)
lossFunc = nn.MSELoss()
optim = torch.optim.Adam(model.parameters(), lr=0.001)
X = torch.tensor(X[:, [1,2,3,4,5,6,9,10,11]], dtype=torch.float32)
Y = torch.tensor(Y[:, 1:], dtype=torch.float32)
label = torch.tensor(label, dtype=torch.float32)
for epoch_i in range(10000):
    optim.zero_grad()
    predY = model(X)
    loss = lossFunc(predY, Y)
    loss.backward()
    optim.step()
    print("epoch: {0}, loss: {1}".format(epoch_i, loss.detach().numpy()))
```

```
MechanicalIndexArray2020 = read2020PredictSample(FLAGS)
MechanicalIndexArray2020 = std_sca.transform(MechanicalIndexArray2020)# 归一化
ManualScoreArray2020pred = model(torch.tensor(MechanicalIndexArray2020[:,
[1,2,3,4,5,6,9,10,11]], dtype=torch.float32))
```

```
neg_pred_idx = set(np.where(np.mean(man_std_sca.inverse_transform(predY.detach().numpy())-
man_std_sca.inverse_transform(Standard_man), 1) < 0)[0])
pos_pred_idx = set(np.arange(Y.shape[0])) - neg_pred_idx
print(" 模型对于训练集的拟合率: ",
(len(pos_pred_idx&pos_man_idx)+len(neg_pred_idx&neg_man_idx))/X.shape[0])
```

```
pred = np.ones((ManualScoreArray2020pred.shape[0], 1))
pred[list(set(np.where(np.mean(man_std_sca.inverse_transform(ManualScoreArray20
20pred.detach().numpy())-man_std_sca.inverse_transform(Standard_man), 1) <
0)[0]))] = 0
```

5. SVM 分类预测代码 (python)

```
svc = SVC()
parameters = [
    {
        'C': [1, 3, 5, 7, 9, 11, 13, 15, 17, 19],
```

```

        'gamma': [0.00001, 0.0001, 0.001, 0.1, 1, 10, 100, 1000],
        'kernel': ['rbf']
    },
    {
        'C': [1, 3, 5, 7, 9, 11, 13, 15, 17, 19],
        'kernel': ['linear']
    }
]
clf = GridSearchCV(svc, parameters, cv=5, n_jobs=8)
clf.fit(X[:, [1,2,3,4,5,6,9,10,11]], label)
print(clf.best_params_)
best_model = clf.best_estimator_

MechanicalIndexArray2020 = read2020PredictSample(FLAG)
MechanicalIndexArray2020 = std_sca.transform(MechanicalIndexArray2020)# 归一化
ManualScoreArray2020pred = best_model.predict(MechanicalIndexArray2020[:,
[0,1,2,3,4,5,8,9,10]])

predY = best_model.predict(X[:, [1,2,3,4,5,6,9,10,11]])
label = label.squeeze()

print(" 模型 对于 训练 集的 拟合 率 : ", 1-
np.where((predY+label)==1)[0].shape[0]/predY.shape[0])

```

6. 混淆矩阵与一致性评判系数（python）

```

def ConfusionMatrix(MecPos, MecNeg, ManPos, ManNeg):
    TP = MecPos & ManPos
    TN = MecNeg & ManNeg
    FP = MecPos & ManNeg
    FN = MecNeg & ManPos
    return TP, TN, FP, FN
class Evaluation():
    def __init__(self, TP, TN, FP, FN):
        super(Evaluation, self).__init__()
        self.TP = TP
        self.TN = TN
        self.FP = FP
        self.FN = FN
        self.P = self.TP / (self.TP + self.FP+np.finfo(np.float32).eps)
        self.R = self.TP / (self.TP + self.FN+np.finfo(np.float32).eps)
        self.F1 = 2*self.P*self.R/(self.P+self.R+np.finfo(np.float32).eps)
        self.oa = (TP + TN) / (TP + FP + TN + FN+np.finfo(np.float32).eps)

```

```

        self.pre = ((TP + FN) * (TP + FP) + (TN + FP) * (TN + FN)) / np.square(TP
+ TN + FP + FN+np.finfo(np.float32).eps)
        self.kappa = (self.oa - self.pre) / (1 - self.pre+np.finfo(np.float32).eps)
        self.iou = TP/(TP+FP+FN+np.finfo(np.float32).eps)
    def Precision(self):
        return self.P
    def Recall(self):
        return self.R
    def F1_score(self):
        return self.F1
    def OA(self):
        return self.oa
    def Kappa(self):
        return self.kappa
    def IoU(self):
        return self.iou
    def evaluation(self):
        return self.F1, self.oa, self.kappa, self.iou

```

7. 问题 3 LSTM 预测模型 (MATLAB)

```

close all
index = 7;
dataone = data8{index};
rmset = [];
for p = 0.6:0.1:0.9
    numTimeStepsTrain = floor(p*size(dataone,1));
    dataTrain = dataone(1:numTimeStepsTrain+1,:);
    dataTest = dataone(numTimeStepsTrain+1:end,:);
    mu = mean(dataTrain,1);
    sig = std(dataTrain,[],1);
    dataTrainStandardized = (dataTrain - mu) ./ sig;
    XTrain = dataTrainStandardized(1:end-1,:);
    YTrain = dataTrainStandardized(2:end,:);
    numFeatures = 3;
    numResponses = 3;
    numHiddenUnits = 100;
    layers = [ ...
        sequenceInputLayer(numFeatures)
        lstmLayer(numHiddenUnits)
        % dropoutLayer(0.6,"Name","drop")
        fullyConnectedLayer(numResponses)
        regressionLayer];
    options = trainingOptions('adam', ...
        'MaxEpochs',200, ...
        'GradientThreshold',1, ...

```

```

        'InitialLearnRate',0.005, ...
        'LearnRateSchedule','piecewise', ...
        'LearnRateDropPeriod',125, ...
        'LearnRateDropFactor',0.2, ...
        'Verbose',0, ...
        'Plots','none');
net = trainNetwork(XTrain,YTrain,layers,options);
dataTestStandardized = (dataTest - mu)./ sig;
XTest = dataTestStandardized(1:end-1,:);

if false
net = predictAndUpdateState(net,XTrain');
[net,YPred] = predictAndUpdateState(net,YTrain(end,:));
numTimeStepsTest = size(XTest,1);
for i = 2:numTimeStepsTest
    [net,YPred(:,i)] = predictAndUpdateState(net,YPred(:,i-1),'ExecutionEnvironment','cpu');
end
else
net = resetState(net);
net = predictAndUpdateState(net,XTrain');
YPred = [];
numTimeStepsTest = size(XTest,1);
for i = 1:numTimeStepsTest
    [net,YPred(:,i)] =
predictAndUpdateState(net,XTest(i,:),'ExecutionEnvironment','cpu');
end
end

YPred = diag(sig)*YPred + mu';
YTest = dataTest(2:end,:);
rmse = sqrt(mean((YPred-YTest).^2,2));
rmseaset = [rmseaset rmse];
end
close all
yyaxis left
yucecishu = (0.6:0.1:0.9)*18/2;
plot(yucecishu,rmseaset(1:2,:))
yyaxis right
plot(yucecishu,rmseaset(3,:))
legend({'带皮硬度预测均方差','去皮硬度均方差','TSS 下降率均方差'},'Location','best')
title('随预测次数增加，预测误差情况')

```