

doi:10.6041/j.issn.1000-1298.2021.11.040

# 苹果货架期 GAN - BP - ANN 预测模型研究

马惠玲<sup>1</sup> 曹梦柯<sup>1</sup> 王 栋<sup>2</sup> 邱凌雨<sup>1</sup> 任小林<sup>2</sup>

(1. 西北农林科技大学生命科学学院, 陕西杨凌 712100; 2. 西北农林科技大学园艺学院, 陕西杨凌 712100)

**摘要:** 准确预测剩余货架期是降低苹果过长贮藏风险的有效途径,目前基于传统动力学模型的预测准确度较低,提出一种基于生成式对抗网络(GAN)改进的反向传播人工神经网络(BP-ANN)苹果货架期预测方法。以0、5、15、25℃下贮藏的“富士”苹果为研究对象,获取果实的12个理化品质指标随贮藏时间变化的取值;分别采用2种特征选择方法对品质指标进行排序,依次累加排序为1~12的品质指标结合贮藏温度作为BP-ANN的输入层变量。通过GAN扩大BP-ANN的训练集样本数量,建立“富士”苹果货架期的GAN-BP-ANN和BP-ANN预测模型。试验结果表明,经过GAN可生成与真实数据分布范围一致的数据集,以真实和生成数据集共同作为训练集构建的GAN-BP-ANN模型其验证集准确度总体高于BP-ANN模型;以稀疏主成分分析(SPCA)选取得到的前1、2、6个品质指标,结合贮藏温度分别作为GAN-BP-ANN模型的输入层对货架期进行预测,其平均相对误差均在0.070以内,决定系数均在0.988以上。

**关键词:** 富士苹果; 品质指标; 货架期预测; 生成式对抗网络; 反向传播人工神经网络

中图分类号: S126; S661.1

文献标识码: A

文章编号: 1000-1298(2021)11-0367-09

OSID:



## Study on Shelf-life Prediction of Apple with GAN - BP - ANN Model

MA Huiling<sup>1</sup> CAO Mengke<sup>1</sup> WANG Dong<sup>2</sup> QIU Lingyu<sup>1</sup> REN Xiaolin<sup>2</sup>

(1. College of Life Science, Northwest A&amp;F University, Yangling, Shaanxi 712100, China

2. College of Horticulture, Northwest A&amp;F University, Yangling, Shaanxi 712100, China)

**Abstract:** Accurately predict the shelf-life of apple is urgently needed in practice. A feasible and non-equipment-depended data collection and model construction method was explored for shelf-life prediction of apple based on quality attributes observations and storage temperature. ‘Fuji’ apples were stored at four different temperatures of 0℃, 5℃, 15℃ and 25℃, respectively. The firmness, soluble solids content, titratable acid, SSC - TA rate, reducing ascorbic acid, starch content, weight loss and color values ( $L$ ,  $a$ ,  $b$ ,  $\Delta E$ ,  $C$ ) were measured periodically to obtain data set of 12 quality features at each storage stage and temperature. Feature selection method of SPCA and ReliefF was used to rank the quality attributes, respectively. Generative adversarial networks (GAN) - back propagation artificial neural network (BP - ANN), and BP - ANN were used to construct regression models between quality feature, storage temperature and shelf-life. Ratio of training set to test set was 3:1. Totally 12 quality attributes were ranked in different orders by different feature selection methods. Using each accumulative combination of 1 ~ 12 th quality attributes and storage temperature as input variables of GAN - BP - ANN and BP - ANN respectively, error rate of the validation set as evaluation criterion of prediction model. The accuracy of the models constructed by feature selection methods of SPCA were higher than that of ReliefF. The accuracy of the models established by GAN - BP - ANN were generally higher than that of the BP - ANN. It showed that GAN can effectively reduce the overfitting of BP - ANN model. Using three selected feature combinations as input variables, respectively, BP - ANN reached an accuracy above 0.930. GAN added BP - ANN can be a novel approach for accurately predict the shelf-life of postharvest ‘Fuji’ apples by using the selected quality attributes and temperature.

**Key words:** ‘Fuji’ apple; quality attributes; shelf-life prediction; generative adversarial networks; back propagation artificial neural network

收稿日期: 2020-12-13 修回日期: 2021-01-22

基金项目: 陕西省科技重大专项(2020zdx03-05-01)和财政部和农业农村部:国家现代农业产业技术体系项目

作者简介: 马惠玲(1965—),女,教授,博士生导师,主要从事果实采后生理与技术研究,E-mail: hl65@nwfau.edu.cn

## 0 引言

我国苹果年产量约占世界总产量的50%<sup>[1]</sup>。准确预测货架期能有效减少贮藏损失,保障上市果实品质。在苹果保鲜技术的研究中,通常测定失重率、可溶性固形物含量、可滴定酸含量、硬度、色值、抗坏血酸含量等品质指标来衡量保鲜效果<sup>[2-5]</sup>。因此,有望从中筛选特征指标或组合来用于采后苹果的货架期预测。目前在一些果蔬上已成功运用品质指标来构建货架期预测模型<sup>[6-8]</sup>。

广泛应用于货架期预测的模型主要包括传统的动力学模型以及机器学习模型等。前人对苹果采后品质指标变化也进行了探究,以及建立了 Arrhenius 货架期预测方程,但预测准确率较低<sup>[9]</sup>,说明常用的动力学模型难以表征苹果采后品质综合变化的复杂性,从而使其预测误差率较高。机器学习模型如反向传播人工神经网络(Back propagation-artificial neural networks, BP-ANN)已经广泛应用于农产品<sup>[10-12]</sup>的货架期预测问题上。然而,由于机器学习模型的复杂性,小样本数据集容易造成过拟合、欠拟合现象。扩充数据集,可以在一定程度上降低过拟合现象,提高模型预测的准确率。SMOTE(Synthetic minority over-sampling technique)及其改进算法的数据合成方法属于线性插值法,其合成的数据和实际数据相关性不强<sup>[13]</sup>。生成式对抗网络(GAN)模型是一种基于深度学习的数据生成方法,其基于博弈理论的数据生成原理,并采用无监督的学习方法,能自动对数据集进行学习,生成高质量的数据<sup>[14]</sup>。目前 GAN 已经应用于在一些小规模样本生成问题上<sup>[13,15-17]</sup>,暂未发现其用于果蔬理化指标数据的生成。

“富士”作为苹果中的优良品种,在我国和日本的种植比例都高达50%以上<sup>[18]</sup>。本文使用 GAN 提升 BP-ANN 模型的预测性能,基于苹果贮藏期间的真实数据来生成采后“富士”苹果的理化品质指标、贮藏温度及货架期数据,扩大 BP-ANN 的训练样本集数量,结合不同的变量排序方式对品质指标进行排序,分别建立品质指标和贮藏温度作为输入变量的货架期预测模型。

## 1 材料与方 法

### 1.1 材料与处理

分别于2016—2018年10月10—11日在陕西省渭南市白水县某果园采收达到商品成熟期的套袋栽培的“富士”苹果(采收前7d脱去最后一层半透明果袋),选取果形规整、表面光洁、大小均匀的套袋果,去果袋,单果套发泡网,运回西北农林科技大

学(3h)。于0℃冷库预冷24h,散去田间热量,再分别在温度0、5、15、25℃和相对湿度85%~90%的条件下贮藏,所有果实均装于加有厚度11μm塑料内衬袋的果框内,袋口松散折叠以保湿。0℃贮藏果前72d每12d取样一次,后198d每24d取样一次;5℃贮藏果前63d每9d取样一次,后108d每18d取样一次;15℃果每6d取样一次,25℃果每3d一次。每次取样随机抽取12个果实,以4个为一组建立3个生物学重复,用于在贮藏期间的各品质指标测定。

### 1.2 货架期苹果理化品质测定

#### 1.2.1 颜色参数

使用白板校准后的 CR-400 型色差计(日本 Konica Minolta 公司),分别用果实赤道线均匀的5个点来测定果实的颜色参数(亮度  $L$ 、红绿度  $a$ 、黄蓝度  $b$ )。总色差  $\Delta E$  和饱和度  $C$  计算公式为

$$\Delta E = \sqrt{(L - L_0)^2 + (a - a_0)^2 + (b - b_0)^2} \quad (1)$$

$$C = \sqrt{a^2 + b^2} \quad (2)$$

式中  $L_0$ 、 $a_0$ 、 $b_0$ ——颜色参数初始值

#### 1.2.2 硬度

沿苹果果实赤道线的阴阳两面各取两点削去1cm×1cm果皮,然后采用GY-3型果蔬硬度计(意大利 Aldo Brue 公司)测定硬度,单位为 N/cm<sup>2</sup>。

#### 1.2.3 可溶性固形物含量、可滴定酸含量和固酸比

沿果实的赤道面,随机选取3个点,每个点去皮后各取10g果肉,用榨汁机榨出汁,用吸管吸取3滴果汁,使用SW-LB32T型折光仪测定苹果的可溶性固形物含量,用质量分数表示。取剩余的果汁,采用酸碱滴定法测定可滴定酸含量<sup>[19]</sup>,用质量分数表示。固酸比为可溶性固形物与可滴定酸质量分数的比值。

#### 1.2.4 抗坏血酸含量

随机称取苹果鲜样5g,采用钼蓝比色法<sup>[20]</sup>测定果实中还原型抗坏血酸的质量比,单位为 mg/kg。

#### 1.2.5 淀粉含量

随机称取苹果鲜样1.5g,采用硫酸蒽酮法<sup>[21]</sup>测定果实中淀粉质量比,单位为 g/kg。

#### 1.2.6 质量损失率

每次取样时称取苹果的鲜质量,以质量随时间下降的百分比计算质量损失率。

### 1.3 货架期苹果感官品质的观测与评定

请10名经过专业培训的人员,对每次取样的12个果实进行品尝和感官品质评分,参照文献[22]的评定项目和权重,每批果实的得分值为10个品尝员评分的平均值。

## 1.4 数据处理

### 1.4.1 理化品质指标排序

#### (1) 稀疏主成分分析

稀疏主成分分析是在主成分分析的基础上引入带有稀疏度的惩罚系数或者不同的系数约束条件,使得到的部分载荷向量为零,从而得到稀疏的主成分<sup>[23]</sup>。

#### (2) ReliefF 算法

ReliefF 算法最早由文献[24]提出,最早用于解决二分类问题,ReliefF 算法是公认的效果较好的 Filter 式特征评估算法。其关键思想是根据属性的值对实例的区分程度去估计这个特征区分邻近样本的能力,特征选择的思路是选取一个特征子集,使得特征子集上的分类错误率最小<sup>[25-26]</sup>。预测的品质指标权重通常取决于最近邻的数量,在本研究中,一共有 12 个品质指标,将最近邻分别设置为从 2 到 11 的数,计算预测权重平均值来作为最终结果,其权重即代表品质属性的重要程度,权重越大,则该品质属性越重要。

### 1.4.2 附加 GAN 的 BP-ANN 货架期预测模型

生成式对抗网络(GAN)是由文献[27]提出的基于博弈论的生成式深度学习算法。GAN 的一般结构如图 1 所示,主要由生成器和判别器两部分组成。由生成器接收随机噪声数据,真实数据的标签为 0,生成数据的标签为 1,由判别器判别是真实数据还是生成器生成的数据,判别器的损失函数为一个二分类模型,可通过交叉熵计算目标函数,其损失函数为

$$J(D) = -\frac{1}{2}E_{x \sim P_{data}(x)} \ln D(x) - \frac{1}{2}E_{z \sim P_z(z)} \ln(1 - D(G(z))) \quad (3)$$

式中  $G, D$ ——生成器和判别器的可微函数

$E$ ——目标函数的期望值

$x$ ——真实样本数据

$z$ ——随机噪声

$G(z)$ ——判别器的生成数据

下角标  $x \sim P_{data}(x)$  表示  $x$  采样于真实数据分布  $P_{data}(x)$ ,  $z \sim P_z(z)$  表示  $z$  采样于真实数据分布  $P_z(z)$ 。

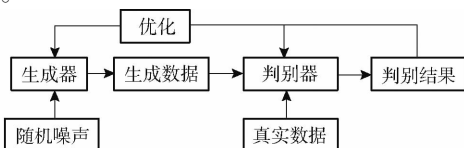


图 1 GAN 网络结构

Fig. 1 Structure of GAN

第 1 项代表  $D$  判断出  $x$  是真实数据的情况,第 2 项代表  $D$  判断数据是否是由生成器  $G$  将噪声矢量  $z$  映射而成的生成数据。 $G$  和  $D$  进行二元零和博弈,GAN 算法的流程为先固定生成器优化判别器,使得判别器的判别准确率最大化,然后固定判别器,优化生成器使得判别器的判别准确率最小。当且仅当  $P_{data}$ (判别真实数据的准确率)等于  $P_g$ (判别生成数据的准确率)时达到全局最优解。

生成器和判别器均采用全连接网络结构。生成器是由两层感知机组成,其输入为随机噪声,第 1 层其激活函数为 ReLU,由 25 个隐藏层神经元组成,第 2 层其激活函数为 Linear,输出维度为 14 的数据。判别器也是由两层感知机组成,其输入为真实数据和生成器生成的假数据,第 1 层由 25 个神经元组成,其激活函数为 ReLU,第 2 层由 1 个神经元组成,其激活函数为 Sigmoid。每训练 3 000 次,保存一次模型。文献[15]通过直接观察生成的菌菇表型图像数据来选择 GAN 生成的图像;文献[13]根据鸭蛋的蛋形指数来选择 GAN 生成的数据;文献[17]根据模型预测的准确率来判断 GAN 模型对于长短期记忆网络(LSTM)的改进作用,从而选择 GAN 生成的数据。在本研究中,苹果的品质指标取值均随着贮藏时间的变化而变化,而且品质指标无法像图像一样可以直接观察,因此通过 GAN 的判别器判别真实数据和生成数据的准确率对生成数据进行初次选择,并通过绘图的形式,将生成数据和真实数据的取值范围进行比较,再通过 GAN-BP-ANN 模型预测货架期的准确性来再次判断 GAN 生成数据的质量。试验平台为 Windows 10 系统,8 GB 内存,500GB SSD,1TB HD, Intel Core i5-5200U, 2.20 GHz, Nvidia GeForce 930M, 2 GB。算法采用 Tensorflow V1.1 GPU 框架和 Python 3.7 实现。

BP-ANN 模型是一种误差反向传播的模型,其通常由输入层、隐藏层和输出层 3 层组成。在此研究中,将品质属性和贮藏温度作为输入层,货架期作为输出层。将得到的品质指标,按照其排序结果,从 1 到 12 逐一叠加,再组合贮藏温度,作为输入变量分别建立扩充数据集的 GAN-BP-ANN 和未扩充数据集的 BP-ANN 货架期预测模型。从图 2 可以看出,基于 GAN 改进的 BP-ANN 模型结构将 GAN 生成的最优解和真实测定的数据同时作为 BP-ANN 的输入层,BP-ANN 模型第 1 层的激活函数为 Tansig,第 2 层的激活函数为 Purelin,若在最大迭代次数内未达到训练目标,则根据相应规则对参数进行优化,如果达到训练目标,则保存模型,并根据保存的模型对验证集进行预测。训练目标为

0.000 1,学习率为 0.01,最大迭代次数为 1 000。由于其初始化的权值和阈值对网络的性能具有较大的影响,在此研究中通过多次建模来选择最优的权值和阈值。

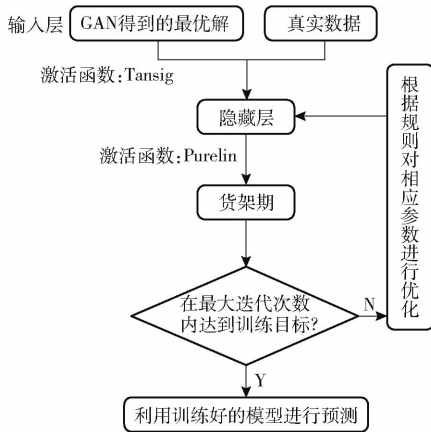


图2 GAN改进的BP-ANN模型

Fig. 2 Optimized BP-ANN model by using GAN

采用平均相对误差和决定系数( $R^2$ )作为模型准确性的评价标准,建模过程通过 Matlab 2019a 软件实现,在进行建模前,将所有数据进行归一化处理。

## 2 结果与分析

### 2.1 不同温度下苹果品质指标的变化趋势

考虑到生产上采用冷藏、气调贮藏苹果的温度通常在  $0^{\circ}\text{C}$ ,采用自然低温贮藏时在  $5\sim 15^{\circ}\text{C}$ ,消费者在室温存放又常为  $20\sim 25^{\circ}\text{C}$ ,因此,苹果采后可能存在的环境温度为  $0\sim 25^{\circ}\text{C}$  范围内的任意

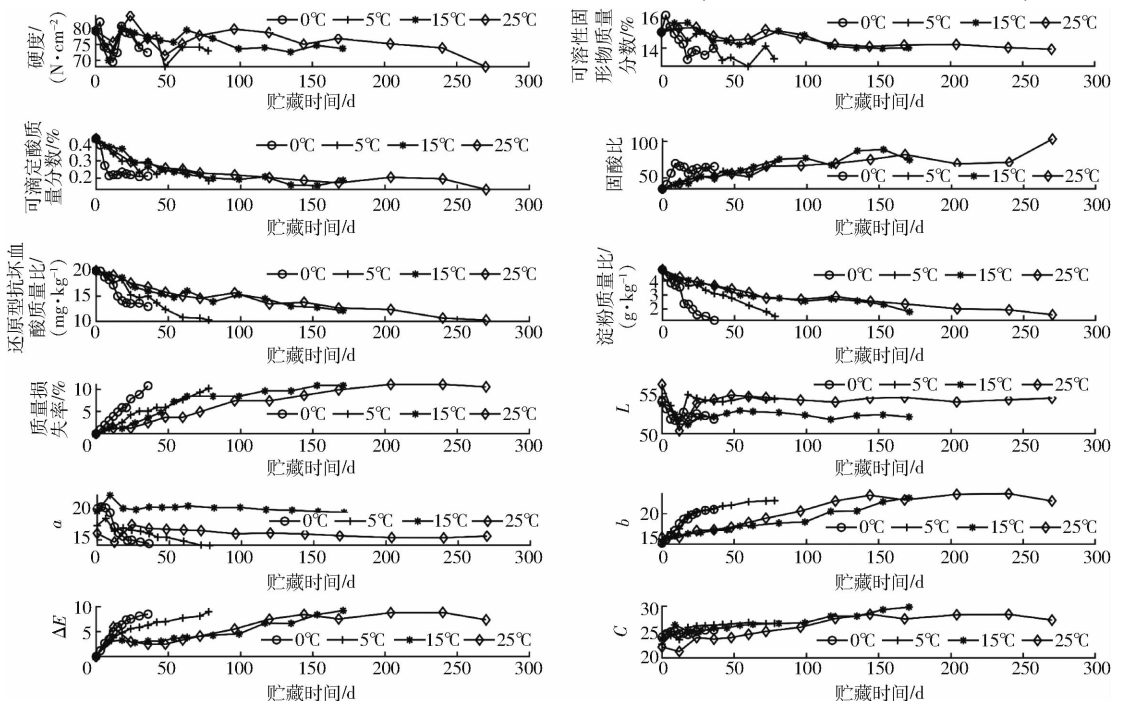


图3 苹果在不同温度下贮藏的品质指标变化曲线

Fig. 3 Change curves of quality attributes of 'Fuji' apples stored at different temperatures

温度,本研究选定其中的 4 种特征温度进行仿真试验。

从图 3 可看出, $0^{\circ}\text{C}$ 下贮藏的苹果 12 种品质指标变化速率最慢,随着温度的升高,变化速率加快。各指标在不同温度下的变化总趋势一致,表现为 3 种类型:①渐降型。首先是可溶性固形物含量,各组可溶性固形物含量在第 2 个观测点均出现短暂上升,这是果实采后初期淀粉等多糖降解量较大,可溶性糖的积累大于消耗,使其总量增加所致。而采后可溶性糖主要是作为呼吸底物而被消耗<sup>[28]</sup>,因此,贮藏(货架)全程呈现总体下降趋势。其次,可滴定酸和还原型抗坏血酸含量也全程下降。抗坏血酸作为一种抗氧化剂,能清除机体内活性氧,延缓衰老<sup>[29-30]</sup>,其中性和碱性环境下极易被氧化,它和可滴定酸总量的含量逐渐减少意味着苹果果实贮藏过程中除了酸味会变淡外,果实营养品质也在下降;淀粉作为果实细胞重要的贮藏性物质,其降解与果实的软化有关<sup>[31]</sup>。②上升型。质量损失率呈现上升趋势,导致果实质量损失的原因主要是水分的丧失和呼吸消耗<sup>[32]</sup>。随着贮藏温度的升高和贮藏时间的延长,黄蓝度  $b$ 、总色差  $\Delta E$ 、色彩饱和度  $C$  总体呈现上升状态,主要是由于在贮藏期间叶绿素逐渐被降解<sup>[33]</sup>,反映了苹果的底色逐渐黄化,光泽变暗。③起伏+渐变型。硬度在前 7 个观测点和  $a$ 、 $L$  在前 3~5 个观测点取值均呈起伏式变化,以后缓慢下降,虽然随着贮藏时间的延长,果胶物质逐渐被细胞壁酶降解,细胞壁结构逐渐发生变化,果实的硬度总

体下降<sup>[32]</sup>,但是,不同苹果采后硬度下降快慢不同,“富士”苹果以其“宁烂不绵”而著称,此文结果从数据上展示了该品种硬度在贮藏全程下降慢的属性;5~25℃组的前3个观测点均表现  $a$  先上升再下降,对应地, $L$  先下降再上升,可见,苹果采后在货架期红度短暂增大,亮度短暂下降,这与果实采后后熟有关<sup>[33]</sup>。“富士”苹果果实的这种变化特性在其它研究中也有所报道<sup>[29,34]</sup>。0℃组可溶性固形物含量和  $a$  的初期变化幅度小, $L$  却也先急降后上升,表明除了与红度有关外, $L$  还受到果面其它属性,如果粉厚度等影响,是与糖分、红绿度不完全相关的独立属性。

### 2.2 品质指标重要性排序结果

在建立货架期预测模型时,用简化的数据集作为模型的输入变量可以节省运算时间和预算。稀疏主成分分析 (SPCA) 和 ReliefF 算法对品质指标进行排序的标准不同,SPCA 算法是将高维数据向低维子空间映射降维,ReliefF 算法侧重于自变量对因变量的区分程度。表 1 列出这 2 种分析方法对品质属性的排序结果,可以看出,两种排序结果有很大区别,SPCA 中,当  $k=1$  时,得到的排序第一的品质指标为质量损失率。ReliefF 算法中,与货架期关联度最大的为  $L$ 。

表 1 2 种特征提取方法得到的品质指标排序结果

Tab.1 Ranking of quality attributes obtained by three feature extraction methods

排序	稀疏主成分分析 (SPCA)		ReliefF 算法	
	品质指标	$k$	品质指标	权重
1	质量损失率	1	$L$	0.058
2	$b$	2	还原型抗坏血酸质量比	0.055
3	淀粉质量比	3	$a$	0.050
4	可滴定酸质量分数	4	$C$	0.030
5	还原型抗坏血酸质量比	5	$\Delta E$	0.028
6	$\Delta E$	6	可滴定酸质量分数	0.018
7	可溶性固形物质量分数	7	硬度	0.015
8	固酸比	8	$b$	0.004
9	硬度	9	固酸比	-0.009
10	$C$	10	淀粉质量比	-0.014
11	$L$	11	可溶性固形物质量分数	-0.014
12	$a$	12	质量损失率	-0.020

### 2.3 附加 GAN 的 BP - ANN 货架期预测模型构建

#### 2.3.1 GAN 生成的有效数据

在构建预测模型时,数据集越大,预测模型能学习到的特征越多并且越准确,从而越能避免模型的过拟合和欠拟合现象,使得模型的预测准确性越高。而实际上,经常由于试验材料、经费等众多问题使试验获得的数据集较小。生成式对抗网

络 (GAN) 是一种深度学习算法,其通过生成器和判别器互相博弈来提高生成数据的准确性。随着迭代次数的增加,GAN 的生成器会生成不同的数据。在此研究中,为了使品质属性的取值更加可靠,将年份之间作为重复,获得的理化指标和对应货架期的平均值数据为 51 组,判别器对此真实数据和生成数据判别准确率如图 4 所示。可以看出,随着迭代次数的增加,判别准确率逐渐接近于 0.50,当迭代次数为 33 000 次时,判别真实数据的准确率为 0.51,判别生成数据的准确率也为 0.51,均最接近于 0.50,继续增大迭代次数至  $2.0 \times 10^5$ ,判别器对两组数据判别准确率偏差增大,说明迭代次数为 33 000 时的生成器和判别器之间已经接近纳什均衡,即生成数据与真实数据已经非常相似。由图 5 (图中参数序号 1~14 分别表示硬度 ( $N/cm^2$ )、可溶性固形物质量分数 (%)、可滴定酸质量分数 (%)、固酸比、还原型抗坏血酸质量比 ( $mg/kg$ )、淀粉质量比 ( $g/kg$ )、质量损失率 (%)、 $L$ 、 $a$ 、 $b$ 、 $\Delta E$ 、 $C$ 、贮藏温度、货架期)显示,生成数据各指标取值均在真实数据取值范围之内,直观表现了二者的相似性。由于前人的研究均没有对 GAN 生成数据的合理量有具体约定<sup>[13-15]</sup>,故选择迭代次数 33 000 次时生成的 38 组数据作为通过 GAN 进行数据生成的结果,用于后续的模型构建。

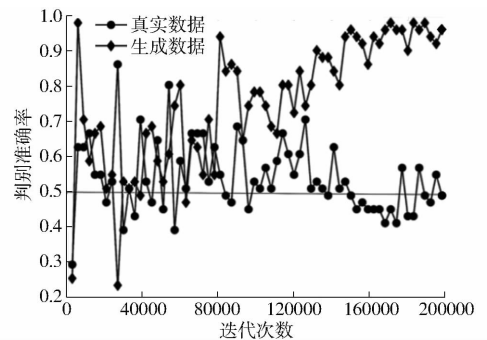


图 4 生成式对抗网络判别器的判别准确率

Fig. 4 Discriminant accuracy of GAN discriminator

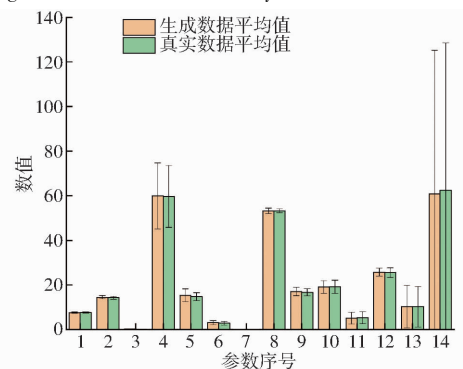


图 5 真实数据和 GAN 生成的有效数据比较

Fig. 5 Comparison of real data and valid data generated by GAN

按训练集和验证集为3:1的比例进行建模和验证,训练集经生成式对抗网络(GAN)扩充得到的38组生成数据不用于验证集验证,即分别采用真实数据的3/4(38组)、真实数据+生成数据(76组)作为训练集,选取每个贮藏温度下的剩余的1/4组数据(其中25℃共4组数据,其它3个温度各3组),共13组真实数据作为验证集来构建BP-ANN和GAN-BP-ANN货架期预测模型。

### 2.3.2 GAN对ReliefF-BP-ANN预测货架期准确性的影响

为了评估GAN对于BP-ANN货架期预测模型的改进作用,分别按照ReliefF排序方法的结果,将1~12个品质指标依次累加,再加上贮藏温度作为输入层变量,分别建立经过训练集扩充的ReliefF-GAN-BP-ANN模型和未经训练集扩充的ReliefF-BP-ANN模型。为了有效评估GAN的作用,均将各个模型重复训练100次,取其平均值作为最后的结果,模型的最大训练次数、学习率和激活函数均相同。由表2可知,ReliefF-GAN-BP-ANN模型其训练集平均相对误差在0~0.095之间,ReliefF-BP-ANN模型其训练集平均相对误差在0~0.112之间,ReliefF-GAN-BP-ANN相比ReliefF-BP-ANN模型,其建模效率没有明显的改善。由图6可直观地看出,采用ReliefF-GAN-BP-ANN模型其验证集的平均相对误差均低于ReliefF-BP-ANN模型,决定系数均高于ReliefF-BP-ANN模型,表明附加GAN的BP-ANN模型(ReliefF-GAN-BP-ANN)对本研究中建模较单独采用ReliefF-BP-ANN的准确度有明显提高,有效地增

表2 采用ReliefF排序方法构建GAN-BP-ANN和BP-ANN模型训练集的平均相对误差

Tab.2 Mean relative error of GAN-BP-ANN and BP-ANN model training set based on ReliefF

品质指 标累加 数量	GAN-BP-ANN			BP-ANN		
	建模时 长/s	迭代 次数	平均相 对误差	建模时 长/s	迭代 次数	平均相 对误差
1	3.038	1000	0.003	3.004	1000	0.110
2	0.455	80	0.018	0.392	42	0.057
3	0.277	16	0.023	0.354	12	0.111
4	0.313	7	0.022	0.308	11	0.100
5	0.289	5	0	0.407	8	0.112
6	0.240	6	0.084	0.355	8	0.108
7	0.291	4	0.038	0.367	7	0.112
8	0.237	4	0.006	0.349	9	0.055
9	0.301	6	0.017	0.303	9	0.066
10	0.320	6	0.012	0.330	4	0.010
11	0.291	7	0.095	0.365	5	0.079
12	0.324	7	0.006	0.380	7	0.011

加了模型的预测准确率,说明GAN生成数据的有效性及增大数据集改进BP-ANN预测货架期的准确性。由图6可知,用ReliefF对品质指标进行排序后,分别构建ReliefF-BP-ANN和ReliefF-GAN-BP-ANN模型,当采用排序前8的8个品质指标即L、还原型抗坏血酸含量、a、C、 $\Delta E$ 、可滴定酸含量、硬度和贮藏温度作为ReliefF-GAN-BP-ANN的输入变量时,验证集中13组数据的预测值与真实值的平均相对误差最小,为0.154,决定系数为0.957。

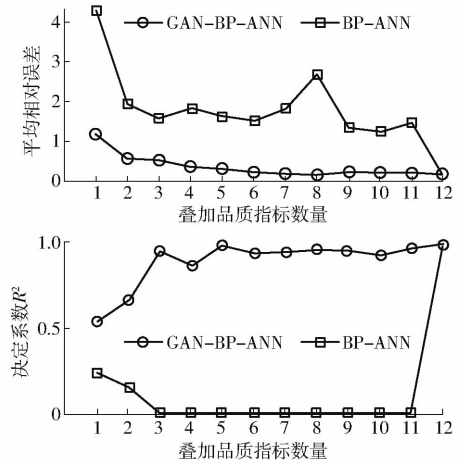


图6 采用ReliefF排序方法构建的GAN-BP-ANN和BP-ANN模型验证集的平均相对误差及决定系数  
Fig.6 Mean relative error, determination coefficient of GAN-BP-ANN and BP-ANN model validation set based on ReliefF

### 2.3.3 GAN对SPCA-BP-ANN预测货架期准确性的影响

按照SPCA排序方法的结果,将1~12个品质指标依次累加,再加上贮藏温度作为输入层变量,分别建立经过训练集扩充的SPCA-GAN-BP-ANN模型和未经训练集扩充的SPCA-BP-ANN模型。均将各个模型重复训练100次,取其平均值作为最后的结果,模型的最大训练次数、学习率和激活函数均相同。由表3可知,SPCA-GAN-BP-ANN模型其训练集平均相对误差在0~0.018之间,SPCA-BP-ANN模型其训练集平均相对误差在0~0.019之间,均小于ReliefF-GAN-BP-ANN和ReliefF-BP-ANN模型,SPCA-GAN-BP-ANN相比SPCA-BP-ANN模型,其建模效率没有明显的改善。由图7可见,采用SPCA-GAN-BP-ANN模型其验证集的平均相对误差也均低于SPCA-BP-ANN模型,决定系数也均高于BP-ANN模型。当采用排序第1的品质指标即质量损失率和贮藏温度(D1组)一起作为GAN-BP-ANN的输入变量时,验证集中13组数据的预测值与真实值的平均相对误差最小,为0.052,决定系数为0.989。当采用排

表 3 SPCA 排序方法构建 GAN-BP-ANN 和 BP-ANN 模型训练集的平均相对误差

Tab.3 Mean relative error of GAN-BP-ANN and BP-ANN model training set based on SPCA

品质指标累加数量	GAN-BP-ANN			BP-ANN		
	建模时长/s	迭代次数	平均相对误差	建模时长/s	迭代次数	平均相对误差
1	0.239	4	0.007	0.291	5	0.002
2	0.232	4	0.004	0.322	5	0.003
3	0.320	7	0.012	0.411	6	0.004
4	0.231	5	0.004	0.398	7	0
5	0.315	6	0.008	0.384	6	0.002
6	0.238	3	0.013	0.258	4	0.008
7	0.340	4	0.006	0.361	6	0.017
8	0.309	5	0.001	0.323	6	0.004
9	0.269	6	0.002	0.341	9	0.019
10	0.225	5	0.012	0.349	5	0.012
11	0.291	4	0.018	0.322	9	0.014
12	0.269	5	0.006	0.355	9	0.011

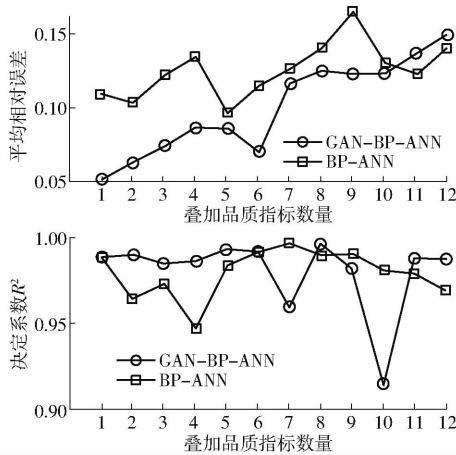


图 7 采用 SPCA 排序方法构建的 GAN-BP-ANN 和 BP-ANN 模型验证集的平均相对误差及决定系数

Fig.7 Mean relative error, determination coefficient of GAN-BP-ANN and BP-ANN model validation set based on SPCA

序前 2 的两个品质指标和贮藏温度作为输入层时,即质量损失率、*b* 和贮藏温度 (D2 组) 作为 GAN-BP-ANN 的输入变量时,验证集中 13 组数据的预

表 4 多元线性回归 (MLR)、决策树 (DT) 和支持向量机 (SVM) 货架期预测模型训练集和验证集的平均相对误差和决定系数

Tab.4 Mean relative error and determination coefficient of training set and validation set of multiple linear regression (MLR), decision tree (DT) and support vector machine (SVM) shelf-life prediction models

参数	MLR		DT		SVM	
	训练集	验证集	训练集	验证集	训练集	验证集
D1 组平均相对误差	4.506	1.662	0.460	1.025	4.283 1	1.714 6
D1 组决定系数	0.727	0.830	0.828	0.745	0.796 8	0.861 8
D2 组平均相对误差	1.001	1.000	0.427	0.883	3.575 8	1.347 9
D2 组决定系数	0.807	0.877	0.882	0.492	0.785 8	0.889 3
D3 组平均相对误差	0.994 7	0.997 6	0.740	0.602	2.965 8	0.952 5
D3 组决定系数	0.887 9	0.905	0.880	0.287	0.867 6	0.897 2

测值与真实值的平均相对误差为 0.064, 决定系数为 0.990。当采用排序前 6 的 6 个品质指标即质量损失率、*b*、淀粉含量、可滴定酸含量、还原型抗坏血酸含量、 $\Delta E$  和贮藏温度 (D3 组) 作为 GAN-BP-ANN 的输入变量时,验证集中 13 组数据的预测值与真实值的平均相对误差为 0.070, 决定系数为 0.992。综合可知,2 种特征选择方法中,SPCA 通过特征累加所构建的模型其验证集平均相对误差最小,通过 SPCA 挑选出 3 组 GAN-BP-ANN 建模的特征品质指标,即:质量损失率和贮藏温度 (D1 组);质量损失率、*b* 和贮藏温度 (D2 组);质量损失率、*b*、淀粉含量、可滴定酸含量、还原型抗坏血酸含量、 $\Delta E$  和贮藏温度 (D3 组),其验证集中 13 组数据的相对误差分别为 0.052、0.064 和 0.070, 低于 BP-ANN 的 0.109、0.104 和 0.115, 即附加 GAN 的模型把预测准确度从平均 0.891 提高到 0.938, 提高了 0.047; 验证集中预测值和真实值决定系数  $R^2$  分别为 0.989、0.990 和 0.992, 大于等于未附加 GAN 模型的 0.989、0.963 和 0.991。这说明输入层为 D1、D2 和 D3 组所构建的 GAN-BP-ANN 模型可以较为准确地预测采后苹果的货架期。

2.3.4 GAN-BP-ANN 模型与其它货架期预测模型比较

常用的货架期预测模型还包括多元线性回归 (MLR)、决策树 (DT) 和支持向量机 (SVM) 等。分别建立品质属性与货架期之间的多元线性回归 (MLR)、决策树 (DT) 模型和支持向量机 (SVM) 等, 使用上述通过 SPCA 得到的最优模型的输入变量作为输入变量, 即分别使用质量损失率和贮藏温度 (D1 组);质量损失率、*b* 和贮藏温度 (D2 组);质量损失率、*b*、淀粉含量、可滴定酸含量、还原型抗坏血酸含量、 $\Delta E$  和贮藏温度 (D3 组) 作为输入变量, 其训练集和验证集平均相对误差和决定系数如表 4 所示, 可以看出其训练集和验证集的平均相对误差均比较高, 决定系数均比较低。上述选出的最优模型其性能均优于 MLR、DT 和 SVM。

### 3 结论

(1) 采用 GAN 法对观测数据集进行扩充, 迭代次数 33 000 次时生成器和判别器之间接近纳什均衡, 生成数据均在真实数据的分布范围之内。

(2) 2 种变量排序法下, 均以附加 GAN 的 BP-ANN 所建模型对货架期的预测准确度高。且以

SPCA 法排序结果构建 GAN-BP-ANN 模型的平均相对误差较 ReliefF 更低, 部分验证集的平均相对误差均在 0.07 以内, 比未附加 GAN 的 BP-ANN 模型预测准确度提升了 0.047。结合 SPCA 法特征变量选择的 GAN-BP-ANN 模型被确定为预测苹果货架期的有效方法。

### 参 考 文 献

- [1] LI M R, GUO J Q, HE J Q, et al. Possible impact of climate change on apple yield in Northwest China[J]. *Theoretical and Applied Climatology*, 2020, 139(1): 191-203.
- [2] 孙志栋, 田方, 张仁杰, 等. 1-MCP 和简易气调处理对采后子芋贮藏品质的影响[J]. *热带作物学报*, 2016, 37(7): 1424-1429.  
SUN Zhidong, TIAN Fang, ZHANG Renjie, et al. Effects of 1-MCP and simply controlled atmosphere treatment on postharvest son-taro quality in storage period[J]. *Chinese Journal of Tropical Crops*, 2016, 37(7): 1424-1429. (in Chinese)
- [3] 胡位歆, 刘东红, 丁甜, 等. 翠冠梨采后货架期预测模型的构建[J]. *中国食品学报*, 2017, 17(7): 119-128.  
HU Weixin, LIU Donghong, DING Tian, et al. Building shelf-life prediction model of postharvest 'Cuiguang' pear[J]. *Journal of Chinese Institute of Food Science and Technology*, 2017, 17(7): 119-128. (in Chinese)
- [4] 杨玉荣, 赵晓敏, 李芸, 等. 1-MCP 处理对采后库尔勒香梨货架期品质及生理的影响[J]. *食品科技*, 2015, 40(5): 37-41.  
YANG Yurong, ZHAO Xiaomin, LI Yun, et al. Effect of 1-MCP treatment on qualities of Korla fragrant pear fruit during shelf-life[J]. *Food Science and Technology*, 2015, 40(5): 37-41. (in Chinese)
- [5] 祝美云, 李小月, 梁丽松, 等. 不同采后处理对'阿巴特'西洋梨果实冷藏后货架期品质的影响[J]. *中国农学通报*, 2015, 31(33): 273-278.  
ZHU Meiyun, LI Xiaoyue, LIANG Lisong, et al. Effect of different postharvest treatments on quality of 'Abate' pear during shelf life after cold storage[J]. *Chinese Agricultural Science Bulletin*, 2015, 31(33): 273-278. (in Chinese)
- [6] 谢晶, 张利平, 苏辉, 等. 上海青蔬菜的品质变化动力学模型及货架期预测[J]. *农业工程学报*, 2013, 29(15): 271-278.  
XIE Jing, ZHANG Liping, SU Hui, et al. Quality kinetic model and shelf life prediction of green vegetable (*Brassica rapa* var. *chinensis*) [J]. *Transactions of the CSAE*, 2013, 29(15): 271-278. (in Chinese)
- [7] 刘春菊, 钱旻, 宋江峰, 等. 速冻莲藕片贮藏过程中品质变化动力学模型[J]. *农业工程学报*, 2017, 33(6): 301-308.  
LIU Chunju, QIAN Min, SONG Jiangfeng, et al. Establishment of dynamic model for quality change in frozen lotus root slices during storage[J]. *Transactions of the CSAE*, 2017, 33(6): 301-308. (in Chinese)
- [8] 胡位歆, 丁甜, 刘东红. 草莓采后货架期预测模型的构建[J]. *中国食品学报*, 2017, 17(3): 186-193.  
HU Weiqin, DING Tian, LIU Donghong. Building shelf-life prediction models of postharvest strawberry[J]. *Journal of Chinese Institute of Food Science and Technology*, 2017, 17(3): 186-193. (in Chinese)
- [9] 史卫娜, 王若琳, 马惠玲, 等. 基于品质变化对温度响应规律的苹果贮藏寿命预测[J]. *北方园艺*, 2016(15): 116-123.  
SHI Weina, WANG Ruolin, MA Huiling, et al. Shelf-life prediction of apple based on quality responses to temperature[J]. *Northern Horticulture*, 2016(15): 116-123. (in Chinese)
- [10] 刘雪, 李亚妹, 刘娇, 等. 基于 BP 神经网络的鲜鸡蛋货架期预测模型[J/OL]. *农业机械学报*, 2015, 46(10): 328-334.  
LIU Xue, LI Yamei, LIU Jiao, et al. BP neural network based prediction model for fresh egg's shelf life [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(10): 328-334. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20151044&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20151044&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2015.10.044. (in Chinese)
- [11] 丁捷, 唐艳, 刘春燕, 等. 速冻青稞鱼面储运过程中货架期预测模型构建[J]. *中国粮油学报*, 2019, 34(2): 26-33.  
DING Jie, TANG Yan, LIU Chunyan, et al. Construction of prediction model for the shelf-life of quick-frozen highland barley fish noodles[J]. *Journal of the Chinese Cereals and oils Association*, 2019, 34(2): 26-33. (in Chinese)
- [12] 潘治利, 黄忠民, 王娜, 等. BP 神经网络结合有效积温预测速冻水饺变温冷藏货架期[J]. *农业工程学报*, 2012, 28(22): 284-289.  
PAN Zhili, HUANG Zhongmin, WANG Na, et al. Prediction of shelf life for quick-frozen dumpling based on BP neural network and effective accumulated temperature theory[J]. *Transactions of the CSAE*, 2012, 284-289. (in Chinese)
- [13] 李庆旭, 王巧华, 马美湖. 基于生成式对抗网络的禽蛋图像数据生成研究[J/OL]. *农业机械学报*, 2021, 52(2): 236-245.  
LI Qingxu, WANG Qiaohua, MA Meihu. Generating poultry egg image data based on generative adversarial network [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*. 2021, 52(2): 236-245. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20210222&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20210222&flag=1). DOI:10.6041/j.issn.1000-1298.2021.02.022. (in Chinese)
- [14] 柴梦婷, 朱远平. 生成式对抗网络研究与应用进展[J]. *计算机工程*, 2019, 45(9): 222-234.  
CHAI Mengting, ZHU Yuanping. Research and application progress of generative adversarial networks [J]. *Computer Engineering*, 2019, 45(9): 222-234. (in Chinese)
- [15] 袁培森, 吴茂盛, 翟肇裕, 等. 基于 GAN 网络的菌菇表型数据生成研究[J/OL]. *农业机械学报*, 2019, 50(12): 231-239.



- YUANG Peisen, ZHU Maosheng, ZHAI Zhaoyu, et al. Mushroom phenotypic generation based on generative adversarial network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 231-239. [http://www.jcsam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20191226&journal\\_id=jcsam](http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20191226&journal_id=jcsam). DOI: 10.6041/j.issn.1000-1298.2019.12.026. (in Chinese)
- [16] 张鹏强, 刘冰, 余旭初, 等. 面向高光谱影像分类的生成式对抗网络[J]. 测绘通报, 2020(3): 29-34.  
ZHANG Pengqiang, LIU Bing, YU Xuchu, et al. Generative adversarial networks for hyperspectral image classification[J]. Acta Geodaetica et Cartographica Sinica, 2020(3): 29-34. (in Chinese)
- [17] 蒋华伟, 张磊. 基于长短期记忆生成对抗网络的小麦品质多指标预测模型[J]. 电子与信息学报, 2020, 42(12): 2865-2872.  
JIANG Huawei, ZHANG Lei. Multi-index prediction model of wheat quality based on longshort-term memory and generative adversarial network[J]. Journal of Electronics & Information Technology, 2020, 42(12): 2865-2872. (in Chinese)
- [18] 木生, 田琳. 当前国内外苹果主栽品种的构成与发展趋势[J]. 烟台果树, 2012(1): 37.
- [19] 叶青, 江志波. 酸碱滴定法与碘滴定法测定维生素 C[J]. 理化检验(化学分册), 2007, 43(5): 410.
- [20] 李军. 钼蓝比色法测定还原型维生素 C[J]. 食品科学, 2000, 21(8): 42-45.  
LI Jun. Determination of reduced vitamin C by molybdenum blue colorimetry[J]. Food Science, 2000, 21(8): 42-45. (in Chinese)
- [21] 齐秀东, 魏建梅, 赵美微, 等. ‘京白梨’果实成熟软化与糖、淀粉代谢及其基因表达的关系[J]. 中国农业科学, 2015, 48(13): 2591-2599.  
QI Xiudong, WEI Jianmei, ZHAO Meiwei, et al. Relationship between fruit softening and the metabolism of sugar and starch and their related-gene expression in post-harvest ‘Jingbaili’ fruits[J]. Scientia Agricultura Sinica, 2015, 48(13): 2591-2599. (in Chinese)
- [22] 蔡骋, 李永超, 马惠玲, 等. 基于介电特征选择的苹果内部品质无损分级[J]. 农业工程学报, 2013, 29(21): 279-287.  
CAI Cheng, LI Yongchao, MA Huiling, et al. Nondestructive classification of internal quality of apple based on dielectric feature selection[J]. Transactions of the CASE, 2013, 29(21): 279-287. (in Chinese)
- [23] ZOU H, HASTIE T, TIBSHIRANI R. Sparse principal component analysis[J]. Journal of Computational and Graphical Statistics, 2006, 15(2): 265-286.
- [24] KIRA K, RENDELL L. The feature selection problem; traditional methods and a new algorithm[C] // Tenth National Conference on Artificial Intelligence, 1992: 12.
- [25] 伍杰华. 基于 RReliefF 特征选择算法的复杂网络链接分类[J]. 计算机工程, 2017, 43(8): 214-220.  
WU Jiehua. Complex network link classification based on RReliefF feature selection algorithm[J]. Computer Engineering, 2017, 43(8): 214-220. (in Chinese)
- [26] ROBNIKŠIKONJA M, KONONENKO I. An adaptation of Relief for attribute estimation in regression[C] // Fourteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 1997.
- [27] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C] // Advances in Neural Information Processing Systems, 2014: 2672-2680.
- [28] 付瑞敏, 邢文会, 常慧萍, 等. 生物可食性活性膜对红富士苹果采后防腐保鲜的研究[J]. 食品科技, 2016, 41(2): 48-54.  
FU Ruimin, XING Wenhui, CHANG Huiping, et al. Effect on postharvest storage of Fuji apples with biological edible membrane[J]. Food Science and Technology, 2016, 41(2): 48-54. (in Chinese)
- [29] 彭贞贞, 叶旗慧, 徐晓艳, 等. 1-甲基环丙烯处理对红富士苹果贮藏品质的影响[J]. 浙江大学学报(农业与生命科学版), 2020, 46(1): 83-92.  
PENG Zhenzhen, YE Qihui, XU Xiaoyan, et al. Effect of 1-methylcyclopropene treatment on storage quality of red Fuji apple[J]. Journal of Zhejiang University (Agriculture and Life Science), 2020, 46(1): 83-92. (in Chinese)
- [30] 魏建梅, 齐秀东, 闫芳教. 采后‘嘎拉’苹果果实糖和淀粉代谢及关键酶基因表达特性[J]. 北方园艺, 2015(19): 126-131.  
WEI Jianmei, QI Xiudong, YAN Fangjiao. Sugar and starch metabolism and the key-enzyme gene expression in postharvest ‘Gala’ apple fruit[J]. Northern Horticulture, 2015(19): 126-131. (in Chinese)
- [31] 唐先谱, 李喜宏, 张彪, 等. 复合保鲜剂对鲜切苹果贮藏品质影响的研究[J]. 中国食品添加剂, 2018(2): 138-143.  
TANG Xianpu, LI Xihong, ZHANG Biao, et al. Research on the effect of compound preservatives on the storage quality of fresh cut apple[J]. China Food Additives, 2018(2): 138-143. (in Chinese)
- [32] 张海新, 宁久丽, 及华. 果实采后品质和生理变化研究进展[J]. 河北农业科学, 2010, 14(2): 54-56.  
ZHANG Haixin, NING Jiuli, JI Hua. Research advances on quality and physiological changes of postharvest fruit[J]. Journal of Hebei Agricultural Sciences, 2010, 14(2): 54-56. (in Chinese)
- [33] 李海登, 蒋佳男, 张春萌, 等. 纳米 ZnO 保鲜膜对苹果保鲜效果的研究[J]. 食品科技, 2019, 44(10): 33-36.  
LI Haideng, JIANG Jianan, ZHANG Chunmeng, et al. The preservation effect of nanometer ZnO plastic film on apple[J]. Food Science and Technology, 2019, 44(10): 33-36. (in Chinese)
- [34] 程顺昌, 冷俊颖, 任小林, 等. 不同环丙烯类乙烯抑制剂对苹果常温贮藏保鲜效果的影响[J]. 农业工程学报, 2012, 28(6): 269-273.  
CHENG Shunchang, LENG Junying, REN Xiaolin, et al. Effect of different 1-substituted cyclopropenes as ethylene inhibitors on postharvest physiology of apple stored at ambient temperature[J]. Transactions of the CASE, 2012, 28(6): 269-273. (in Chinese)