

参赛密码 \_\_\_\_\_  
(由组委会填写)

第十一届华为杯全国研究生数学建模竞赛

学 校	安徽大学
参赛队号	10357002
队员姓名	1.胡鑫
	2.马闯
	3.刘梦丹

参赛密码 \_\_\_\_\_  
(由组委会填写)



## 第十一届华为杯全国研究生数学建模竞赛

### 题 目                   **基于多种智能优化算法和营养健康** **的中国果蔬发展战略研究**

#### 摘                   要：

水果和蔬菜对人体的营养健康具有举足轻重的作用。预测我国果蔬的消费与生产趋势，科学规划与调整我国果蔬的中长期种植模式，具有重要的战略意义。

针对问题一，首先利用主成分因子分析对蔬菜和水果的营养指标进行筛选，得到主要营养指标有维生素 A、维生素 B1、维生素 B2、维生素 B3、维生素 B6、维生素 C、维生素 E、铁、钙、锌、膳食纤维和蛋白质。其次，为了筛选出主要水果和蔬菜进行研究，构建 0—1 模糊规划模型，同时，为了求解该模型，提出粒子群优化遗传 (PSO-GA) 算法，选出的主要水果包括苹果、柑橘类、梨、桃、香蕉、葡萄、红枣、猕猴桃、李子和西瓜，选出的主要蔬菜包括大豆、西红柿、大白菜、茄子、菠菜、胡萝卜、青椒、菜花和蘑菇。再次，为了对果蔬的人均消费量进行估计，提出了消费量直接估计模型和基于损耗率的人均消费量估计模型估计 2001 年—2012 年果蔬消费量数据，然后提出灰色预测模型、BP 神经网络预测模型、支持向量机回归模型和基于加权 power 平均 (WPA) 算子的组合预测模型对果蔬的未来生产趋势进行预测，结果表明主要水果均呈现上升趋势，其中苹果、柑橘类和梨消费量增长趋势最为明显，主要蔬菜除大白菜有一定的波动呈下降趋势以外，其它均呈不同程度的增长，其中西红柿和土豆的消费量增长趋势最明显。

针对问题二，根据预测模型预测出未来人口数量，从而计算未来主要果蔬的人均摄入量，将该摄入量与标准摄入量区间进行比较，得到锌、维生素 B6、维生素 B2、维生素 B1、膳食纤维的摄入量低于标准范围，维生素 E 和维生素

A 的这些营养素的摄入量高于标准范围，其他营养素的摄入量趋于正常。为了对中国居民未来人体营养健康状况进行评价，提出了一种连续区间有序加权平均（COWA）算子，并建立总体营养评价价值指标，结果表明，随着时间的推移，中国居民的人体营养健康状况会逐渐恶化，尤其是因为维生素 A 的过多摄入会导致各种健康疾病。

针对问题三，先测算蔬菜和水果中各种营养素的含量，考虑蔬菜和水果之间的相互替代，再利用自组织特征映射（SOM）聚类分析模型对水果和蔬菜进行可替代聚类分析，将水果分为类别一：苹果、梨、桃、猕猴桃、李子；类别二：柑橘、香蕉、葡萄、红枣、西瓜；将蔬菜分为类别一：蘑菇；类别二：土豆、大白菜、菠菜；类别三：西红柿、茄子、胡萝卜、青椒、菜花。每种类别内的果蔬可以看成所含营养素大致一样，可相互替代。因为单位质量的蔬菜中所含营养素要远远高于水果，所以本文不建议水果和蔬菜之间的相互替代。进一步，为了保证营养均衡满足健康需要的条件下合理选择消费产品，依据南方和北方果蔬的不同价格分别建立人均消费量最优化模型，并利用 PSO-GA 算法计算得到不同区域不同季节水果和蔬菜人均消费量及相互替代方案。

针对问题四，考虑居民的购买成本和种植者的收益以及人体摄入的营养均衡，同时兼顾果蔬产品的进出口贸易量和土地面积等因素建立多目标规划模型，并利用 PSO-GA 模型求解未来主要水果和蔬菜的人均消费量，结果表明，应适当减少苹果、柑橘类、梨、西瓜、土豆、西红柿和茄子等的种植面积，加大对桃、香蕉、葡萄、猕猴桃、李子、菠菜、胡萝卜、青椒、菜花的种植。

最后根据模型结果给出了相关的政策建议，并对模型进行了灵敏度分析和优缺点分析。

**关键词：果蔬发展；营养；遗传算法；组合预测；战略**

## 一 问题重述

人体需要的营养素主要有蛋白质、脂肪、维生素、矿物质、糖和水。水果和蔬菜是重要的农产品，主要为人体提供矿物质、维生素、膳食纤维。近年来，中国水果和蔬菜种植面积和产量迅速增长，水果和蔬菜品种也日益丰富，中国居民生活水平不断提高，人们对人体营养均衡的意识也有所增强。然而多数中国居民喜食、饱食、偏食、忽视人体健康所需的营养均衡的传统饮食习惯尚未根本扭转，这就使得我国的果蔬消费（品种和数量）在满足居民身体健康所需均衡营养的意义下，近乎盲目无序，进而影响到果蔬生产。

因此，预测我国果蔬的消费与生产趋势，科学地规划与调整我国果蔬的中长期的种植模式，具有重要的战略意义。本文主要考虑以下几个问题：

一、利用合适的数学模型从附件表格中筛选出主要的水果和蔬菜品种，要求是它们所蕴含营养素无论在成分上还是在含量上都满足研究的需要，同时尝试用多种方法建立数学模型对其消费量进行估计，研究其发展趋势。

二、了解我国居民矿物质、维生素、膳食纤维等营养素摄入现状，结合为保障人体健康所需要的各种营养成分的范围和前面预测的人均消费结果，评价中国居民目前矿物质、维生素、膳食纤维等营养的年摄入水平合理与否。按照水果和蔬菜近期的消费趋势，预测至 2020 年中国居民的人体营养健康状况。

三、不同的蔬菜、水果尽管各种营养素含量各不相同，但营养素的种类大致相近，存在着食用功能的相似性。所以，水果与水果之间、蔬菜与蔬菜之间、水果与蔬菜之间从营养学角度在一定程度上可以相互替代、相互补充。由于每种蔬菜、水果所含有的维生素、矿物质、膳食纤维成分、含量不尽相同，价格也有差异，因而在保证营养均衡满足健康需要条件下，如何选择消费产品是个普遍的问题。如何为当今中国居民提供主要的水果和蔬菜产品的按年度合理人均消费量，使人们能够以较低的购买成本（假定各品种价格按照原有趋势合理变动）满足自身的营养健康需要。

四、为实现人体营养均衡满足健康需要，国家可能需要对水果和蔬菜各品种的生产规模做出战略性调整。一方面国家要考虑到居民人体的营养均衡，并使营养摄入量尽量在合理范围内；另一方面也要顾及居民的购买成本，使其购买成本尽量低；同时还要使种植者能够尽量获得较大收益；而且，作为国家宏观战略，还要考虑进出口贸易、土地面积等其他因素。基于上述因素，建立数学模型重新计算中国居民主要的水果和蔬菜产品的按年度合理人均消费量，并给出到 2020 年我国水果和蔬菜产品生产的调整战略。

五、结合前面的研究结论，给相关部门提供 1000 字左右的政策建议。

## 二 问题分析

对于问题一：

题目要求我们解决两个问题：（1）筛选出主要水果和蔬菜，以便满足研究需要；（2）应用多种方法对消费量进行估计，并研究发展趋势。

对于问题（1），我们首先采用因子分析模型[1]对蔬菜、水果的营养素指标进行筛选，然后以每种水果、蔬菜的日消费量为变量建立最优化模型，应用 PSO-GA 算法[3-6]解出在满足营养需求情况下所需要的主要果蔬种类和日消费量。

对于问题（2），首先由第一问求出的主要果蔬种类和日消费量可以直接估

计出消费量，再根据果蔬的产量和损耗率给出第二种消费量估计方法，最后应用组合预测模型研究其发展趋势。

对于问题二：

题目要求我们摸清居民矿物质、维生素、膳食纤维等营养素摄入现状，并在此基础上结合为保障人体健康所需要的各种营养成分范围和前面预测的人均消费结果，评价中国居民目前矿物质、维生素、膳食纤维等营养的年摄入量水平是否合理。因此我们需要计算各种营养素的人均摄入水平，根据问题一已求出的各种果蔬的消费量，已知人口数，可以直接得到我国营养摄入现状，结合营养素的标准范围，评价是否合理。再根据水果和蔬菜近期的消费趋势，推测至 2020 年中国居民的人体营养健康状况的好坏。由于在问题一中我们已经预测了至 2020 年果蔬的消费量，所以只需预测至 2020 年的人口数量，就可以得到人均消费量，与标准营养素区间进行比较，便可以得到我国人体营养状况的发展趋势。

### 三 符号说明与模型假设

#### 3.1 符号说明

表 1 主要符号说明

符 号	说 明
$PB_i$	各年龄段男性的人口总数
$PG_i$	各年龄段女性的人口总数
$NB_{ij}$	各年龄段男性对第 $j$ 种营养素需求标准值
$NG_{ij}$	各年龄段女性对第 $j$ 种营养素需求标准值
$\overline{NB_j}$	所有男性对第 $j$ 种营养素需求标准平均值为
$\overline{NG_j}$	所有女性对第 $j$ 种营养素需求标准平均值为
$N_j^-$	第 $j$ 种营养素的需求区间的左端点值
$N_j^+$	第 $j$ 种营养素的需求区间的右端点值
$\min\_N_j$	第 $j$ 种营养素的需求最小值
$\max\_N_j$	第 $j$ 种营养素的需求最大值
$c_i$	第 $i$ 中果蔬品种的人均每天消费量
$a_{ij}$	第 $i$ 种果蔬品种第 $j$ 种营养素的含量

$C_i$	第 <i>i</i> 中果蔬的消费量
P	人口数
$O_i$	第 <i>i</i> 中果蔬的产量
$S_1$	果蔬在田间地头到大市场损耗率
$S_2$	大市场到零售市场损耗率
$IM_i$	第 <i>i</i> 中果蔬的进口量
$EX_i$	第 <i>i</i> 中果蔬的出口量
$C_{it}$	在 <i>t</i> 年第 <i>i</i> 种果蔬的年消耗量
$P_t$	在 <i>t</i> 年的人口数
$S_3$	零售市场到餐桌的损耗率
$A_{jt}$	在 <i>t</i> 年第 <i>j</i> 种营养素的人均每日摄入量
$PF_{ik}$	第 <i>i</i> 种水果在第 <i>k</i> 季度的南方的价格
$PF'_{ik}$	第 <i>i</i> 种水果在第 <i>k</i> 季度的北方的价格
$x_{ik}$	为第 <i>i</i> 种水果或蔬菜在第 <i>k</i> 季度的南方消费量
$x'_{ik}$	为第 <i>i</i> 种水果或蔬菜在第 <i>k</i> 季度的北方消费量
$PV_{ik}$	第 <i>i</i> 种蔬菜在第 <i>k</i> 季度的南方的价格
$PV'_{ik}$	第 <i>i</i> 种蔬菜在第 <i>k</i> 季度的北方的价格

### 3.2 模型假设

- (1) 假设生产出来的蔬菜和水果除中间损坏外全部被消费；
- (2) 提供和查询数据真实可靠。
- (3) 各个区域的果蔬的各阶段的损耗率相同；
- (4) 人体对果蔬的摄取量不受个人喜好影响；
- (5) 各种果蔬的价格不相互影响；
- (6) 果蔬市场价位合理。

## 四 模型建立与问题求解

### 4.1 问题一的模型建立与求解：

由于我们需要按年份按区域来划分的各种不同蔬菜和水果的产量（或消费量），但是附件给出的数据有缺失现象，我们主要依据的数据来源包括：国家统计局数据库、农业部种植业管理司数据库、历年中国农业统计资料、历年农产品成本收益资料汇编。有的年份数据缺失，我们采用线性插值的方法进行了补充。

#### 4.1.1 因子分析模型对指标进行筛选

指标筛选原则：指标对人类活动干扰具有明显的响应关系；指标间相互独立、不存在重复信息；能够全面反映人体健康营养需求的不同特征属性。利用主成分因子方法（PCA）分别对果蔬的各个营养素进行统计分析，根据提取主成分个数累计方差超过80%的原则，通过最大方差旋转法（Varimax），经因子载荷矩阵旋转后选择载荷值大于0.6的指标作为下一步待筛选指标适用性分析，利用SPSS软件对蔬菜、水果营养数据进行因子分析。

##### 4.1.1.1 蔬菜营养素指标的筛选

表2 蔬菜营养素指标检验统计量

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.645
Approx. Chi-Square		918.681
Bartlett's Test of Sphericity	df	120
	Sig.	.000

KMO检验统计量是用来比较指标间简单相关系数和偏相关系数的，KMO值越大，说明指标间的相关性越强。KMO值为0.645大于0.5，说明指标间有相关性，可以作因子分析。从表2中还可看到其相应的显著性概率小于0.001，高度显著，说明筛选后的数据适合使用因子分析。

提取主因子：各主因子的方差解释如表3所示：

表3 蔬菜营养素指标累计方差解释情况表

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.732	54.576	54.576	8.732	54.576	54.576	7.396	46.226	46.226
2	1.825	11.404	65.979	1.825	11.404	65.979	2.894	18.089	64.315
3	1.398	8.735	74.714	1.398	8.735	74.714	1.615	10.091	74.406
4	1.174	7.337	82.052	1.174	7.337	82.052	1.223	7.646	82.052
5	.961	6.005	88.057						
6	.747	4.667	92.724						
7	.571	3.567	96.291						
8	.400	2.498	98.789						
9	.139	.867	99.656						
10	.044	.275	99.932						
11	.006	.036	99.968						
12	.003	.016	99.984						

13	.002	.010	99.993						
14	.001	.004	99.998						
15	.000	.002	100.000						
16	4.29E-006	2.686E-005	100.000						

做出各成分特征值碎石图，如图 1：

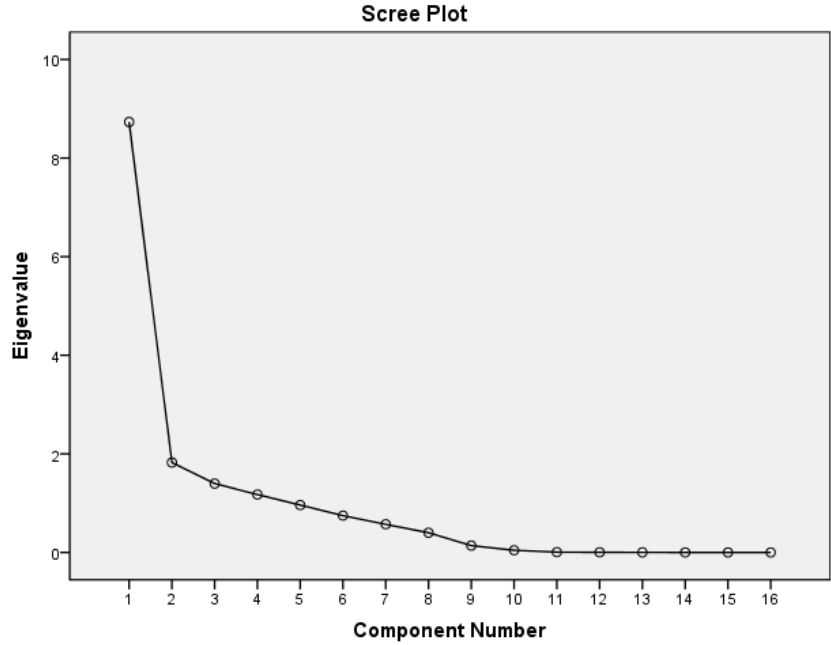


图 1 蔬菜营养指标各成分特征值碎石图

可以看出第 1 个因子特征值最高，对解释原有变量的贡献最大。第 11 个之后的因子特征值都较小，对解释原有变量的贡献均偏小，因此提取 11 个因子是合适的。

旋转后的成分矩阵如表 4 所示：

表 4 旋转后的主成分矩阵

	Component			
	1	2	3	4
维生素B3	.987	-.011	.062	.040
维生素B2	.937	.235	-.014	.044
蛋白质	.931	.336	.028	.032
脂肪	-.251	-.230	-.333	-.542
能量	.872	.464	.054	.038
碳水化合物	.771	.574	.038	.048
膳食纤维	.748	.629	-.036	.062
铁	.183	.912	.043	.008
钙	.306	.757	.429	-.041
钠	.309	.249	.621	.177
维生素A	.132	-.326	.619	-.274

根据累计方差贡献率大于 80%的标准，提取 11 个营养指标构成 4 个主要因



子：

第一主成分：维生素 B3、维生素 B2、蛋白质、能量、碳水化合物、膳食纤维；

第二主成分：铁、钙；

第三主成分：钠、维生素 A；

第四主成分：脂肪。

考虑到在实际饮食中，蛋白质、能量、碳水化合物、脂肪、钠的摄入基本由主食和肉类等获得，因此在这里我们将这些主成分剔除。选择蔬菜的主要营养指标为：维生素 B3、维生素 B2、膳食纤维、铁、钙、钠、维生素 A。

#### 4.1.1.2 水果营养指标筛选

表 5 水果营养指标检验统计量

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.623
Approx. Chi-Square		115.728
Bartlett's Test of Sphericity	df	78
	Sig.	.000

KMO值为0.623大于0.5，说明指标间有相关性，可以用因子分析。且相应的显著性概率小于0.001为高度显著，说明筛选后的数据适合使用因子分析。

提取主因子：各主因子的方差解释如表 6 所示：

表 6 水果营养指标累计方差解释情况表

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.028	23.294	23.294	3.028	23.294	23.294	2.878	22.136	22.136
2	1.856	14.277	37.571	1.856	14.277	37.571	1.782	13.707	35.842
3	1.746	13.429	51.000	1.746	13.429	51.000	1.601	12.317	48.160
4	1.277	9.823	60.824	1.277	9.823	60.824	1.381	10.625	58.785
5	1.026	7.889	68.713	1.026	7.889	68.713	1.291	9.928	68.713
6	.929	7.149	75.862						
7	.800	6.155	82.016						
8	.728	5.597	87.613						
9	.518	3.983	91.597						
10	.389	2.992	94.589						
11	.364	2.798	97.387						
12	.251	1.931	99.318						
13	.089	.682	100.000						

做出各成分特征值碎石图，如图 2

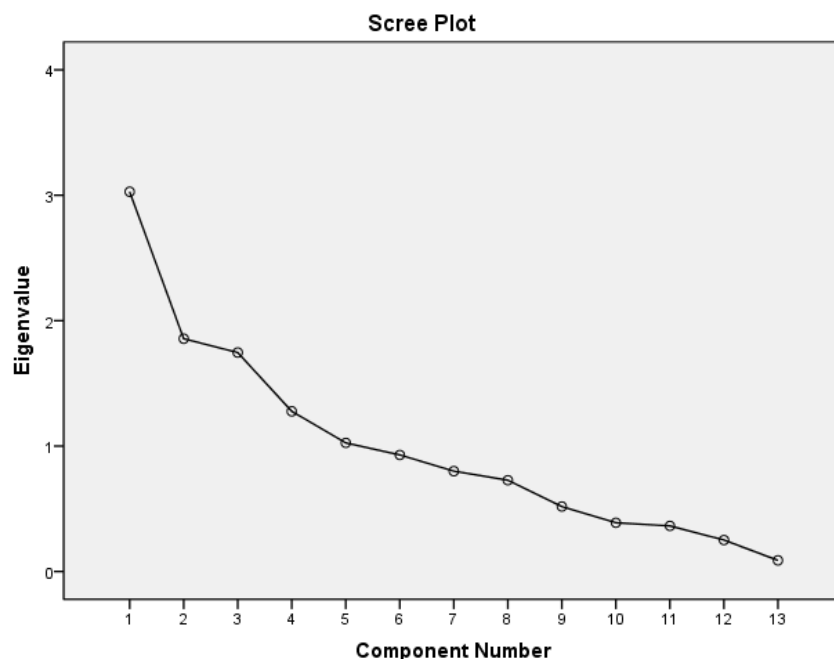


图 2 水果营养素指标各成分特征值碎石图

可以看出第 1 个因子特征值最高，说明其对解释原有变量的贡献最大，后面几个因子特征值都较小，对解释原有变量的贡献很小。

旋转后的成分矩阵如表 7 所示：

表 7 旋转后的主成分矩阵

	Component				
	1	2	3	4	5
蛋白质	.900	-.077	.016	-.240	-.040
膳食纤维	.776	.162	.090	.198	.031
碳水化合物	.693	.464	-.250	.075	-.107
钠	.000	.721	-.234	.157	.067
锌	.634	.169	.139	.300	.308
维生素E	.174	.854	.066	-.157	-.024
维生素B2	-.039	-.069	.894	-.069	.049
维生素B6	.066	-.115	.690	.373	-.148
维生素C	.184	-.079	-.051	.758	.079
铁	.155	-.155	-.280	-.554	.075
维生素B1	-.184	-.061	-.007	.205	.765
维生素A	-.203	-.092	.059	.208	-.735

根据累计方差贡献率大于 80%的标准，提取 12 营养个指标构成 5 个主要因子：

第一主成分：蛋白质、膳食纤维、碳水化合物、锌；

第二主成分：钠、维生素 E；

第三主成分：维生素 B2、维生素 B6；

第四主成分：维生素C、铁；

第五主成分：维生素B1、维生素A。

按选择蔬菜的主要营养指标的方法，可以筛选出水果的主要营养指标为：

膳食纤维、锌、维生素 E、维生素 B2、维生素 B6、维生素 C、铁、维生素 B1、维生素 A。

#### 4.1.1.3 模型结果

结合蔬菜和水果由因子分析筛选得到的结果，可选择营养指标为：维生素 A、维生素 B1、维生素 B2、维生素 B3、维生素 B6、维生素 C、维生素 E、铁、钙、锌、膳食纤维、蛋白质。

#### 4.1.2 模糊规划模型的建立与求解

为了选取主要水果和蔬菜品种，本节先考虑营养素的标准区间，然后建立 0—1 模糊规划模型选取果蔬种类，为了求解该模型，提出粒子群优化遗传（PSO-GA）算法。

##### 4.1.2.1 标准营养素区间数模型建立

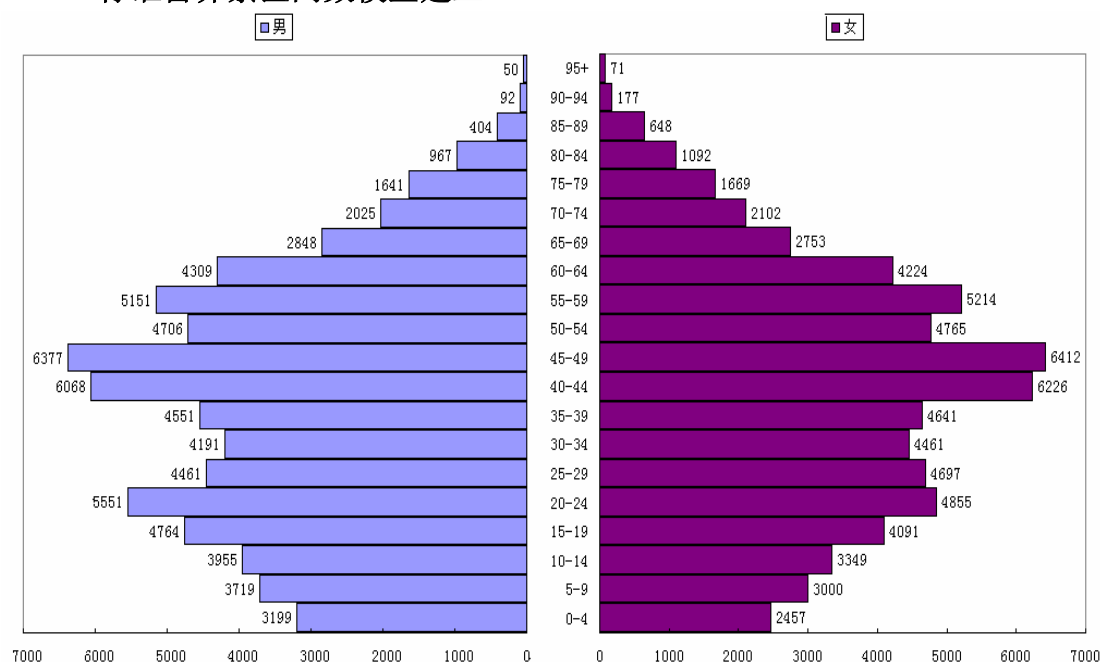


图 3 2013 年末中国人口年龄段结构图（万人）

如图 3 所示，设  $PB_i$ ， $PG_i$  分别表示  $[5i-5, 5i-1](i=1,2,\dots,19)$  年龄区间的男性和女性的人口总数， $PB_{20}$ ， $PG_{20}$  分别表示  $[95, +\infty)$  年龄区间的男性和女性的人口总数。

通过附件 4 计算出第  $j$  种营养素在  $[5i-5, 5i-1](i=1,2,\dots,19), [95, +\infty)$  这 20 个年龄段男性、女性需求的标准量，记为： $NB_{ij}$ ， $NG_{ij}$ 。则男性、女性对第  $j$  种营养素需求标准平均值为：

$$\overline{NB_j} = \frac{\sum_{i=1}^{20} PB_i NB_{ij}}{\sum_{i=1}^{20} PB_i} \quad (1)$$

$$\overline{NG_j} = \frac{\sum_{i=1}^{20} PG_i NG_{ij}}{\sum_{i=1}^{20} PG_i} \quad (2)$$

则第  $j$  种营养素的需求区间为:

$$NI_j = [N_j^-, N_j^+] \quad (3)$$

其中,  $N_j^- = \min\{\overline{NB_j}, \overline{NG_j}\}$ ,  $N_j^+ = \max\{\overline{NB_j}, \overline{NG_j}\}$ 。

计算结果如表 8:

表 8 各种营养素的需求区间

	$N_j^-$	$N_j^+$	$\min\_N_j$	$\max\_N_j$
蛋白质/g	55	65	15	110
膳食纤维/g	20	30	5	60
VA/ug	700	800	400	1200
VB1/mg	1.1	1.3	0.2	1.8
VB2/mg	1.1	1.65	0.5	1.7
VB3/mg	13	19.5	4.5	28
VC/mg	60	100	40	130
VD/ug	5	7.5	5	10
钙/mg	600	750	300	1200
钾/mg	2000	2120	500	2500
锌/mg	13	16	1.5	21.5
铁/mg	11	16	0.3	35
VB6/mg	1.5	2	0.1	4.0

#### 4.1.2.2 区间型隶属函数的确定

定义 1. 第  $j$  种营养素的隶属函数  $u_j(x)$ :

$$u_j(x) = \begin{cases} 0 & , 0 \leq x \leq \min\_N_j \\ \frac{x - \min\_N_j}{N_j^- - \min\_N_j} & , \min\_N_j < x < N_j^- \\ 1 & , N_j^- \leq x \leq N_j^+ \\ \frac{\max\_N_j - N_j^+}{\max\_N_j - x} & , N_j^+ < x < \max\_N_j \\ 0 & , \max\_N_j \leq x \end{cases} \quad (4)$$

当  $u_j(x)=1$  时, 表明该种营养素的量  $x$  满足人体需求标准, 当  $u_j(x)=0$  时,

表明第  $j$  种营养素的量  $x$  不满足人体需求标准，当  $0 < u_j(x) < 1$  时， $u_j(x)$  越大说明这种营养素的量  $x$  越接近人体需求标准。

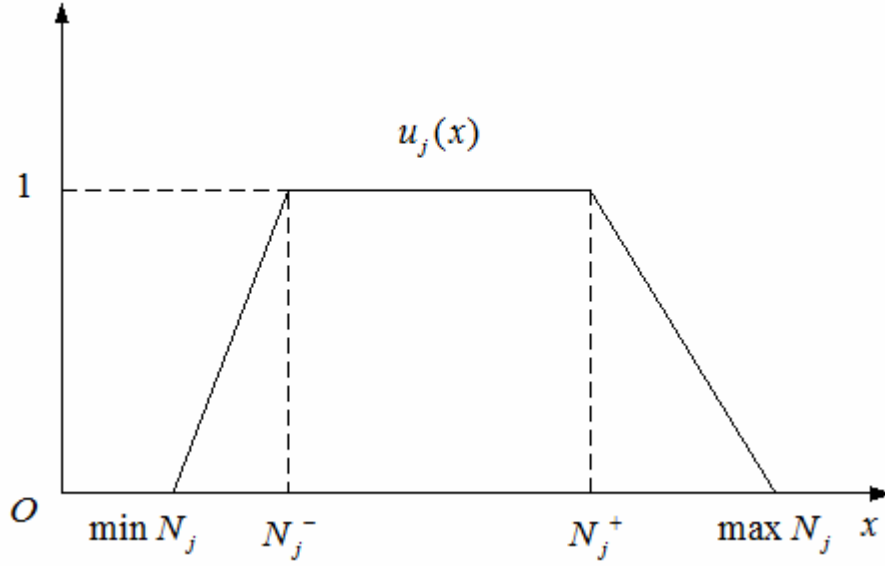


图 4. 隶属度函数

#### 4.1.2.3 0—1 模糊规划模型的建立

设  $c_i$  为第  $i$  中果蔬品种的人均每天消费量，则这种水果是否被筛选为主要品种可定义为：

$$f(c_i) = \begin{cases} 1, & c_i > 0 \\ 0, & c_i = 0 \end{cases}$$

则最优化模型为：

$$\begin{aligned} \min \quad & \sum_i f(c_i) \\ \text{s.t.} \quad & u_j(\sum_i c_i a_{ij}) > \theta \end{aligned} \quad (5)$$

式中  $a_{ij}$  表示第  $i$  种果蔬品种第  $j$  种营养素的含量， $u_j(\sum_i c_i a_{ij})$  表示第  $j$  种营养素的人体摄入量隶属标准摄入量的程度。 $\theta$  为参数，控制营养素的人体摄入量程度，这里设置  $\theta=0.95$ 。

#### 4.1.2.4 PSO-GA 算法求解数学规划

为了求解上述 0—1 模糊规划模型，下面考虑粒子群优化遗传（PSO-GA）算法。

##### (1) 遗传算法

遗传算法（GA）是一种基于自然选择原理和自然遗传机制的搜索（寻优）算法<sup>[8]</sup>，它是模拟自然界中的生命进化机制，在人工系统中实现特定目标的优化。遗传算法的实质是通过群体搜索技术，根据适者生存的原则逐代进化，最终得到最优解或准最优解。它必须做以下操作：初始群体的产生、求每一个体的适应度、

根据适者生存的原则选择优良个体、被选出的优良个体两两配对，通过随机交叉其染色体的基因并随机变异某些染色体的基因后生成下一代群体，按此方法使群体逐代进化，直到满足进化终止条件。

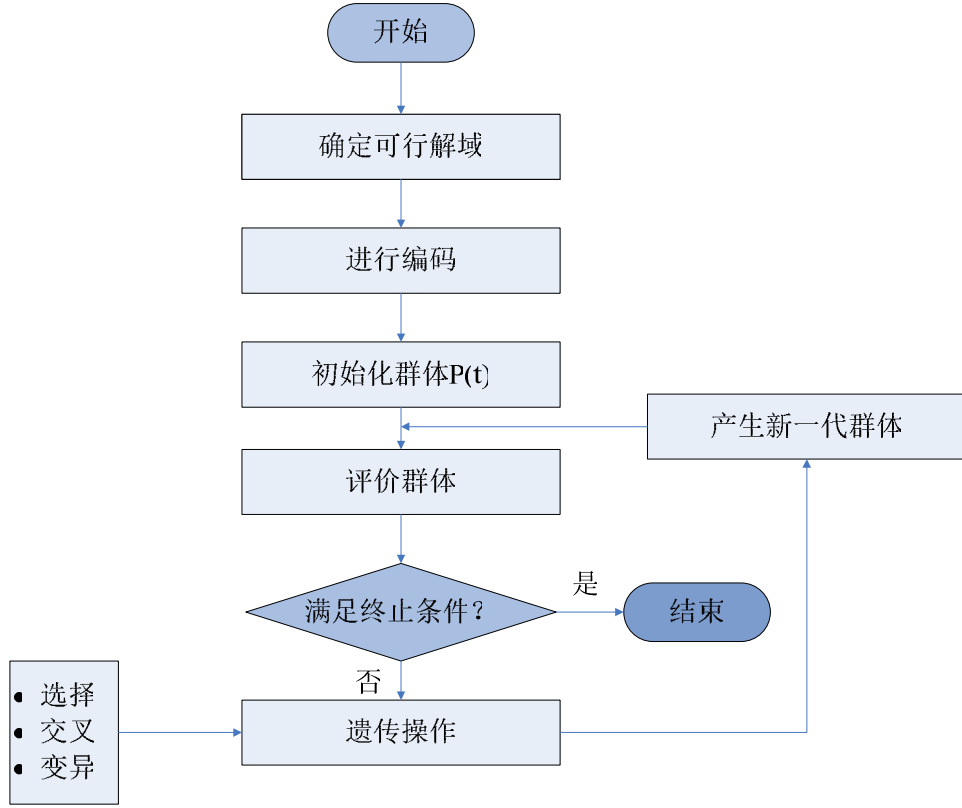


图 5 遗传算法（GA）流程图

其实现步骤如下：

**Step 1** 确定遗传算法中相应的参数，如种群规模  $M$ ，变异概率  $p_m$ ，交叉概率

$p_c$ ，进化代数阈值  $iter$  等；

**Step 2** 随机生成初始种群，即初始化种群；

**Step 3** 计算目标函数值以评价群体；

**Step 4** 若满足停止条件（算法已收敛、超过进化代数等），则算法结束，否则进入 Step 5；

**Step 5** 进行遗传操作，即选择，交叉，变异，生成新一代群体，并回到 Step 3。

## （2）粒子群优化（PSO）算法

粒子群优化算法是基于群体智能理论的优化算法，它是通过群体中粒子间的合作与竞争产生的群体智能知道优化搜索。粒子群优化算法保留了基于种群的全局搜索策略，又采用了一种相对简单的速度-位移模型，避免了复杂的遗传算子操作。同时它特有的记忆功能使其可以动态跟踪当前的搜索情况而调整其搜索策略。

假设  $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$  是微粒  $i$  的当前位置， $V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,n})$  是微粒  $i$  的当前飞行速度，那么，基本粒子群优化算法的进化方程如下：

$$v_{i,j}(t+1) = v_{i,j}(t) + c_1 rand_1() (p_{i,j}(t) - x_{i,j}(t)) + c_2 rand_2() (p_{g,j}(t) - x_{i,j}(t)), \quad (6)$$

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1), \quad (7)$$

$$P_i(t) = (p_{i,1}(t), p_{i,2}(t), \dots, p_{i,n}(t)), \quad (8)$$

$$P_g(t) = (p_{g,1}(t), p_{g,2}(t), \dots, p_{g,n}(t)), \quad (9)$$

式中,  $t$  表示迭代代数;  $P_i(t)$  表示微粒  $i$  迄今为止经历过的历史最好位置;  $P_g(t)$  是当前粒子群搜索到的最好位置 (全局最好位置),  $c_1$ 、 $c_2$  为分别为认知学习因子和社会学习因子, 通常在 0-2 间取值。 $rand_1()$ 、 $rand_2()$  是在 [0,1] 上的两个相互独立的随机数<sup>[9-10]</sup>。

PSO 的基本流程可以描述为如图 6:

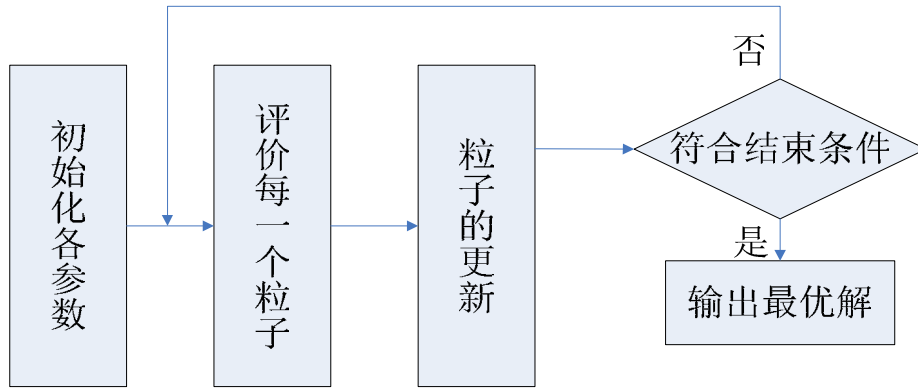


图 6 粒子群 PSO 算法流程图

其实现步骤如下:

**Step 1** 在允许的范围内, 初始化搜索点的位置及其速度, 每个粒子的坐标设置为其当前位置, 且计算出其相应的个体极值 (即个体极值点的适应度值), 而全局极值 (即全局极值点的适应度值) 就是个体极值中最好的, 记录该最优值的粒子序号, 并将  $P_g(t)$  设置为该最优粒子的当前位置;

**Step 2** 计算粒子的适应度值, 即目标函数值, 如果优于该粒子当前的个体极值, 则将  $P_i(t)$  设置为该粒子的位置, 且更新个体极值。如果所有粒子的个体极值中最好的好于当前的全局极值, 则将  $P_g(t)$  设置为该粒子的位置, 记录该粒子的序号, 且更新全局极值;

**Step 3** 根据式 (18) — (21), 对每一个粒子的速度和位置进行更新;

**Step 4** 检验是否符合结束条件, 若当前的迭代次数达到了预先设定的最大次数 (或达到最小错误要求等), 则停止迭代, 输出最优解, 否则转到 Step 2。

### (3) 基于 PSO-GA 方法对上述模型进行求解

文献[11]提出一种名为 PSO-GA 的混合算法, 这种算法以 PSO 和 GA 两种算法为基础, 很好地结合了两者的优点, 同时规避了部分不足, 在优化方程系数方面有更好地表现。

在 PSO-GA 方法目前的研究和使用中，目标值受相关因素的影响可分别建立为线性、指数型、二次型三种模型，其表达式分别如下：

$$Q_1 = \sum_{i=1}^N \tau_i F_i + \tau_0, \quad (10)$$

$$Q_2 = \sum_{i=1}^N \tau_i F_i^{\tau_i} + \tau_0, \quad (11)$$

$$Q_3 = \sum_{i=1}^N \tau_i F_i + \tau_0 + \sum_{i=1}^N \sum_{j=i+1}^N k_{ij} F_i F_j + \sum_{l=1}^N u_l F_l^2, \quad (12)$$

其中  $F_i$  为第  $i$  个影响因素， $\tau_0$ ， $\tau_i$ ， $k_{ij}$  和  $u_i$  为系数， $N$  为影响因素个数。实际操作中，具体流程如图 7：

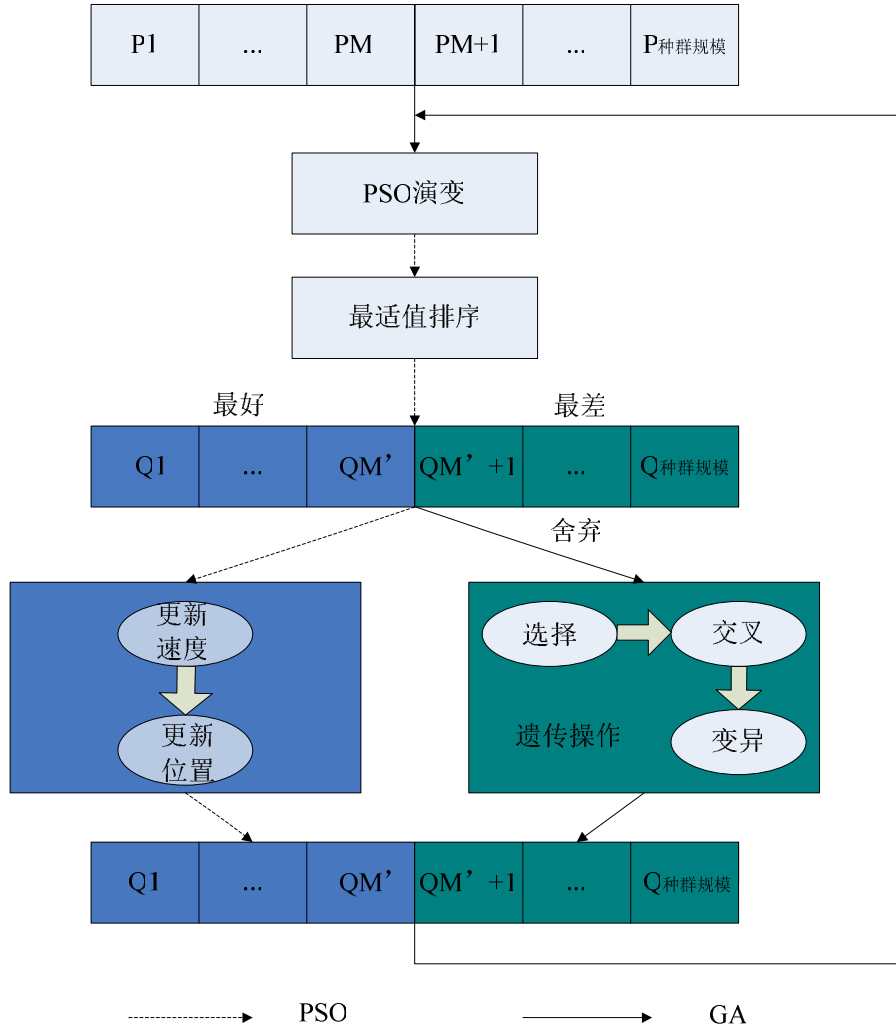


图 7 PSO-GA 算法流程图

其实现步骤如下：

**Step 1** 确定相关系数，如粒子种群规模  $M$ ，由 PSO 演变后保留的粒子规模  $M'$ ，PSO 认知学习因子和社会学系数  $c_1$  及  $c_2$ ，GA 的交叉以及突变概率  $p_c$  和  $p_m$ ，



粒子最大速度  $v_{imax}$ ，PSO 迭代代数阈值  $iter_1$  和 PSO-GA 迭代代数阈值  $iter_2$ ；

**Step 2** 由实际编码确定式 (5) - (7) 的系数；

**Step 3** 若 PSO-GA 迭代达到  $iter_2$ ，则算法结束，输出最优适应度、最优解（即最优位置），否则进入 Step 4；

**Step 4** 若 PSO 迭代达到  $iter_1$ ，则算法结束，否则进入 Step 5；

**Step 5** 类似 PSO 算法，根据下式，更新离子的速度和位置；

$$v_{id}^{t+1} = \chi \cdot [v_{id}^t + c_1 \times rand_1() \times (P_{id}^t - x_{id}^t) + c_2 \times rand_2() \times (P_{gd}^t - x_{id}^t)], \quad (13)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}, \quad (14)$$

其中  $\chi$  是由 Clerc 与 Kennedy 于 2002 年引入的限制因子<sup>[12]</sup>，以保证收敛值与 PSO 方法收敛值一致，其计算方法如式 (4)：

$$\chi = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}, \quad (15)$$

对于  $\varphi > 4$ ，有  $\varphi = c_1 + c_2$ ，一般情况下取  $c_1 = c_2 = 2.05$ ；

**Step 6** 根据适应度对种群粒子进行排序，将适应度最优的  $M'$  个粒子选择，进行式 (7) 和 (8) 操作，更新其速度和位置；

**Step 7** 对于未选择到的粒子，进行 GA 操作，首先根据式 (10)，计算剩余粒子的选择概率，

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j - f_{\max}} = \frac{f_i}{f_{\text{sum}} - f_{\max}}; \quad (16)$$

**Step 8** 根据 Step 1 中  $p_c$  和  $p_m$  以及式 (11) 和 (12)，实施交叉、突变过程，更新粒子群  $M - M'$  的速度和位置：

$$\begin{cases} x_A^{t+1} = \alpha x_B^t + (1 - \alpha) x_A^t \\ x_B^{t+1} = \alpha x_A^t + (1 - \alpha) x_B^t \end{cases}, \quad (17)$$

$$z_i = \begin{cases} v_i, & i \neq h \\ v_h', & i = h \end{cases}; \quad (18)$$

**Step 12** 将 Step 6 和 Step 8 得到的两个粒子群合并，不妨仍记为  $M$ ，回到 Step 3。

#### 4.1.2.5 模型结果

在现实中我们还需要考虑以下几种情况：1) 产量大小。在生活中，某种果蔬的产量高，说明人的消费量大，我们应该把它筛选到主要果蔬中；2) 某种果蔬的大部分营养素含量明显高于所有果蔬平均值，即使其生产量较少，它仍是

人体所需营养素补充的主要果蔬；3) 考虑果蔬产量数据无法查找，我们把这部分剔除到主要果蔬外；4) 为了研究的多样性的需要，某些产量少，人们喜爱的果蔬也应该考虑到主要果蔬中。

应用 PSO-GA 算法求解最优化模型，结合以上几点，我们选择的果蔬分别为：苹果、柑橘类、梨、桃、香蕉、葡萄、红枣、猕猴桃、李子、西瓜、大豆、西红柿、大白菜、茄子、菠菜、胡萝卜、青椒、菜花、蘑菇。其人均消费量如图 8—9 所示

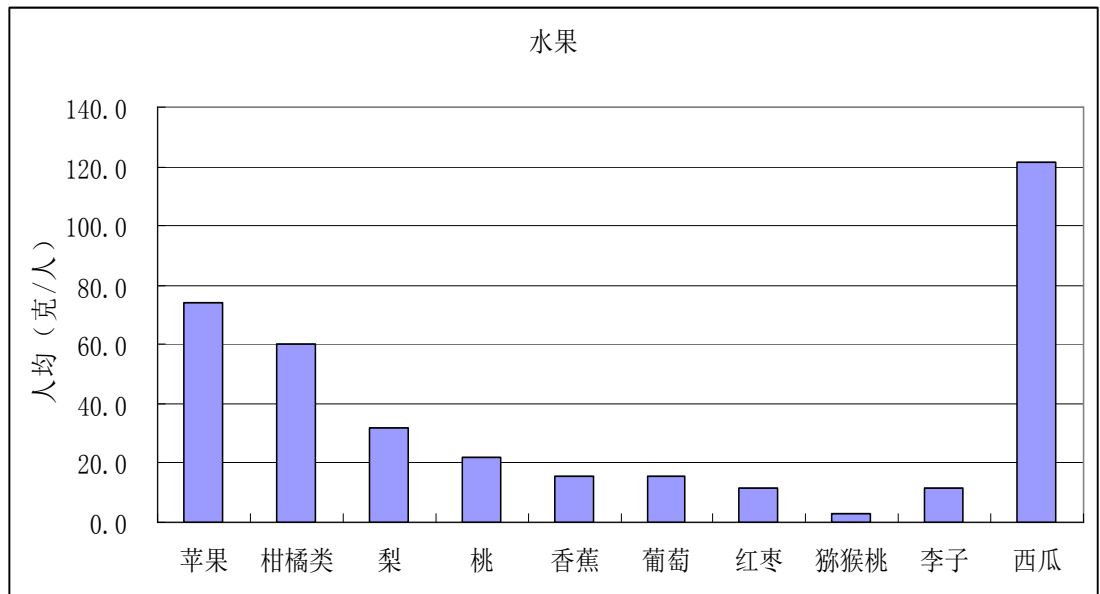


图 8. 各种水果每日的消费量

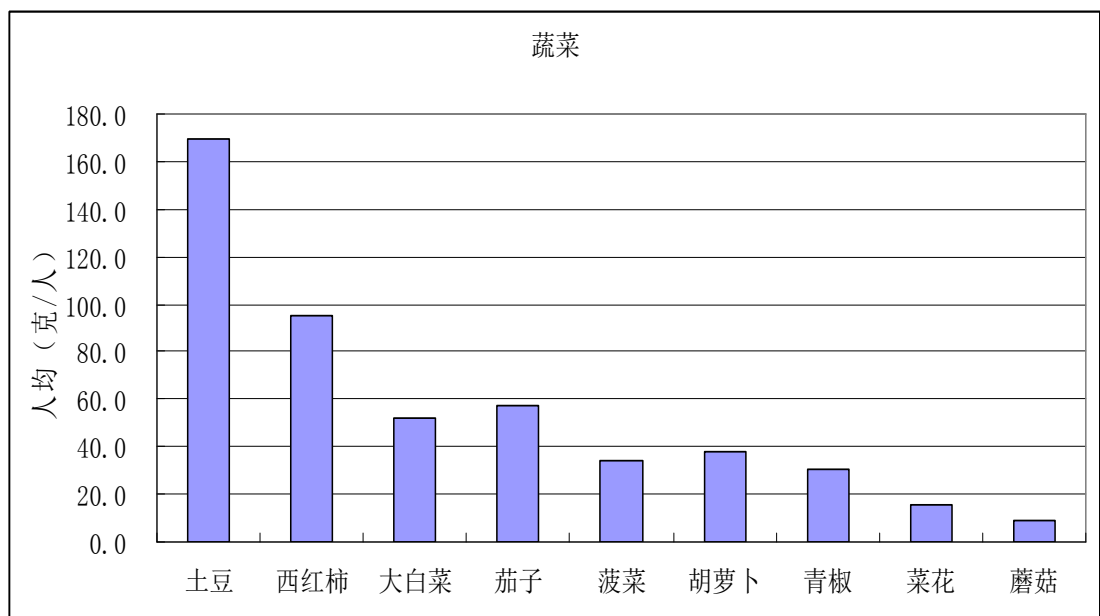


图 9. 各种蔬菜每日的消费量

由图 8 可以看出，人均水果消费量中，西瓜最高，苹果次之，猕猴桃最低；由图 9 可以看出，人均消费量最高的是土豆，其次是西红柿，蘑菇最低。

#### 4.1.3 消费量估计模型一

参照最优化求解的第  $i$  中果蔬品种的人均每天消费量，可以给出一种消费量

估计模型：

$$C_i=365Pc_i \tag{19}$$

式中  $C_i$  为第  $i$  中果蔬的消费量， $P$  为人口数。求解结果如图 10—11 所示：

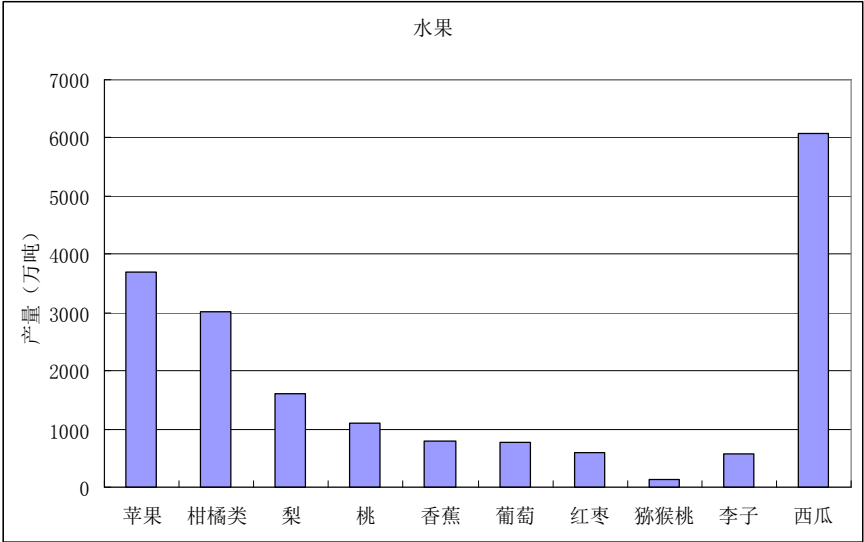


图 10. 各种水果的年消费量

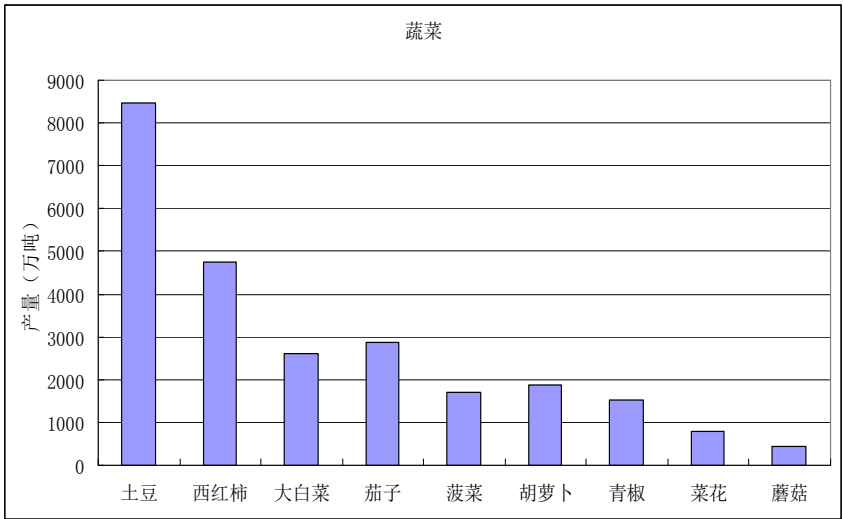


图 11. 各种蔬菜的年消费量

#### 4.1.4 消费量估计模型二

根据查找的果蔬产量数据和附件中的损耗率，我们可以给出一种消费量估计模型：

$$C_i = O_i(1 - S_1)(1 - S_2) + IM_i - EX_i \tag{20}$$

式中  $O_i$  为第  $i$  中果蔬的产量， $S_1$  为果蔬在田间地头到大市场损耗率， $S_2$  为大市场到零售市场损耗率， $IM_i$  为第  $i$  中果蔬的进口量， $EX_i$  为第  $i$  中果蔬的出口量。

计算结果如图 12—13:

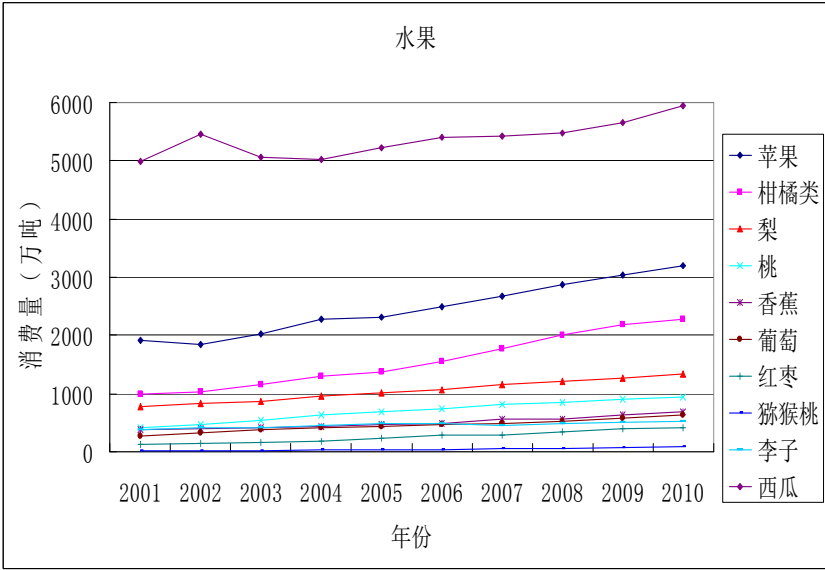


图 12 各种水果的消费量

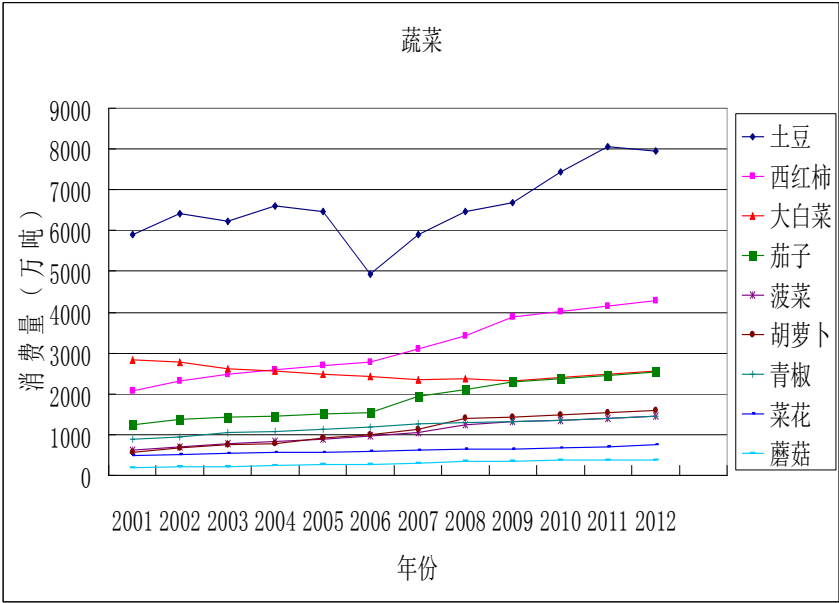


图 13 各种蔬菜的消费量

根据消费量估计模型二计算出的图 12 和图 13 可以看出，水果消费量中，西瓜最高，苹果次之，猕猴桃最低；蔬菜消费量中最高的是土豆，其次是西红柿，蘑菇最低。该结果与消费量估计模型一非常相似。

#### 4.1.5 组合预测模型研究消费量趋势

为了对未来果蔬的消费量进行预测，本节首先提出三种单项预测方法：灰色预测模型、BP 神经网络预测模型和支持向量机回归预测模型，然后提出基于加权 power 平均（WPA）算子的组合预测模型。

##### (1) 灰色预测模型的建立与求解

灰色预测<sup>[7-9]</sup>是灰色系统理论的重要组成部分，其中应用比较广泛的是文献提出的传统 GM (1,1) 模型，它主要适用于预测时间短，数据资料少，波动不

大的系统对象，只需很少的几个数据即可建立模型进行预测。但由于传统 GM (1,1) 模型本身的缺陷，使其仅能适用于短期预测和原始数据序列按指数规律变化且变化速度不是很快的场合。无偏灰色预测模型<sup>[10-11]</sup>消除了传统灰色预测模型本身所固有的偏差，其实只是一种无偏的指数模型，模型准确度优于传统 GM (1,1) 模型。但无论哪一种 GM (1,1) 模型，其预测的几何图形都是一条较为平滑的曲线，对随机波动性大的数据序列进行预测时，预测值起伏不定，影响预测准确度。

首先，建立无偏 GM (1,1) 模型 (unbiased grey model, 简称 UBGM (1,1) 模型)，对原始数据进行趋势化处理。

设原始序列为  $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ ，利用该数据序列建立无偏 GM(1,1)

模型的步骤如下：

**Step1** 作一阶累加形成生成数据序列：

$$x_k^{(1)} = \sum_{i=1}^k x_i^{(0)}, \quad k = 1, 2, \dots, n. \quad (21)$$

**Step2** 确定数据矩阵 B,  $Y_n$ ：

$$B = \begin{bmatrix} -\frac{1}{2}(x_1^{(0)} + x_2^{(1)}) & 1 \\ -\frac{1}{2}(x_2^{(1)} + x_3^{(1)}) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x_{n-1}^{(1)} + x_n^{(1)}) & 1 \end{bmatrix}, \quad Y_n = \begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix} \quad (22)$$

**Step3** 用最小二乘法估计计算一阶线性微分方程的待估计参数  $a$  和  $u$

$$\begin{bmatrix} \hat{a} & \hat{u} \end{bmatrix}^T = (B^T B)^{-1} B^T Y_n. \quad (23)$$

**Step4** 计算无偏 GM (1,1) 模型参数  $b$  和  $A$ . 对呈现指数变化趋势的原始数据序列  $x_k^{(0)} = A e^{b(k-1)}$ ,  $k = 1, 2, \dots, n$ ，作一次累加生成的序列：

$$x_k^{(1)} = \sum_{i=1}^k x_i^{(0)} = A \frac{1 - e^{bk}}{1 - e^b}, \quad k = 1, 2, \dots, n. \quad (24)$$

按传统 GM (1,1) 方法建模可得：

$$\begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} 2(1 - e^b) / (1 + e^b) \\ 2A / (1 + e^b) \end{bmatrix} \quad (25)$$

由上式可求得用传统 GM (1,1) 模型参数  $\hat{a}$ ,  $\hat{u}$  表示的  $b$  和  $A$  的估计为：

$$\hat{b} = \ln \frac{2 - \hat{a}}{2 + \hat{a}}, \quad \hat{A} = \frac{2\hat{u}}{2 + \hat{a}}. \quad (26)$$

**Step5** 建立原始数据序列模型：

$$\hat{x}_1^{(0)} = x_1^{(0)}, \quad \hat{x}_{k+1}^{(0)} = \hat{A}e^{\hat{b}k}, \quad k=1,2,\dots \quad (27)$$

$\hat{x}_{k+1}^{(0)}$ ,  $k=0,1,\dots,n-1$ 为原始数据序列的拟合值,  $\hat{x}_{k+1}^{(0)}$ ,  $k \geq n$ 为原始数据序列的预测值。

根据上述灰色预测模型, 得到蔬菜预测结果如图 14—22 所示。

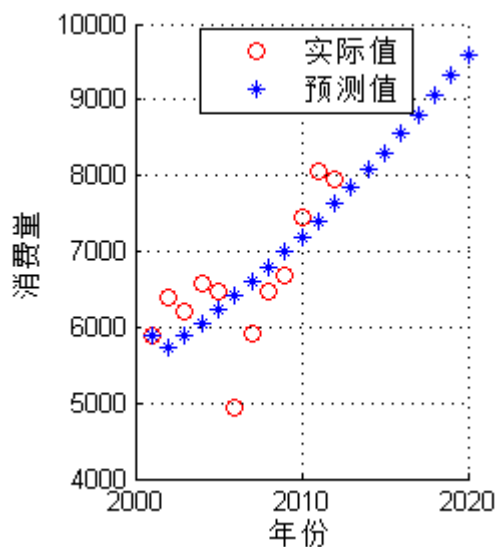


图 14 土豆消费量的实际值与预测值

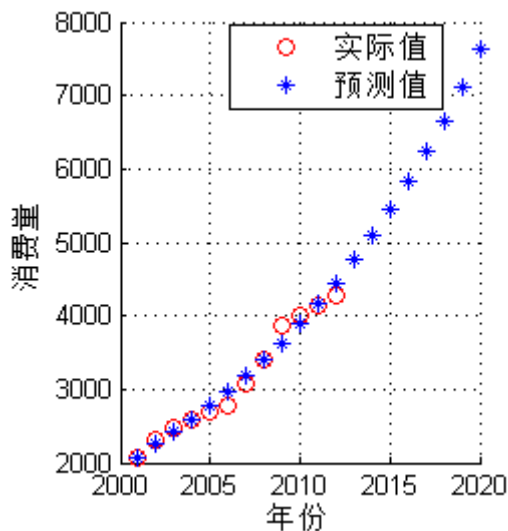


图 15 西红柿消费量的实际值与预测值

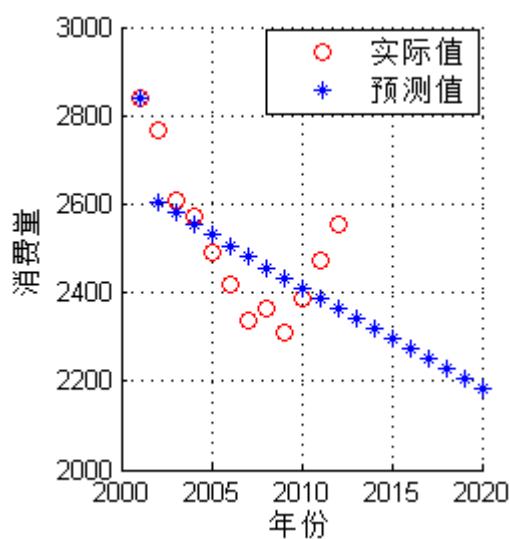


图 16 大白菜消费量的实际值与预测值

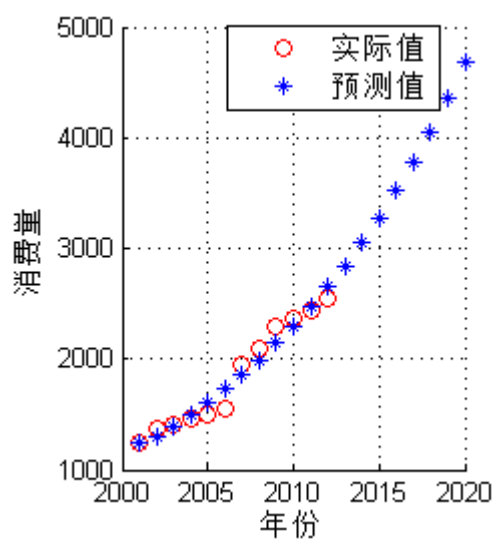


图 17 茄子消费量的实际值与预测值

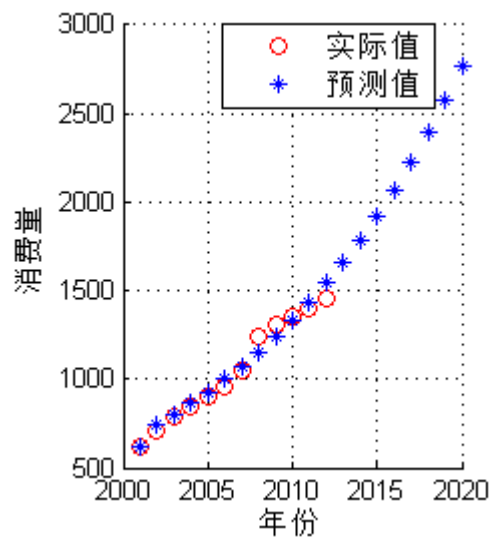


图 18 菠菜消费量的实际值与预测值

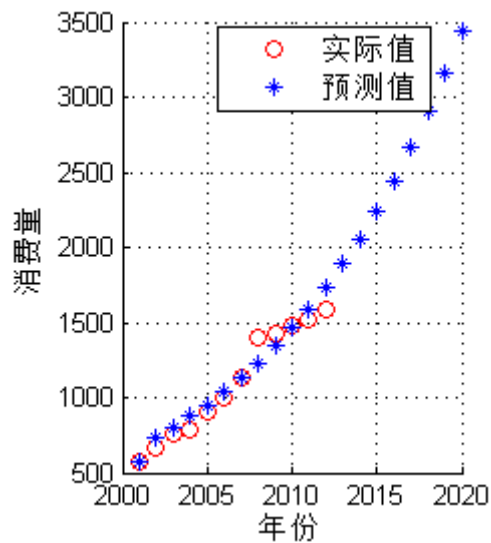


图 19 胡萝卜消费量的实际值与预测

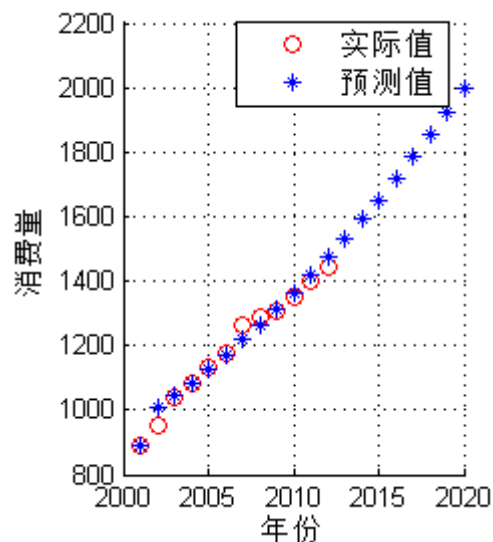


图 20 青椒消费量的实际值与预测值

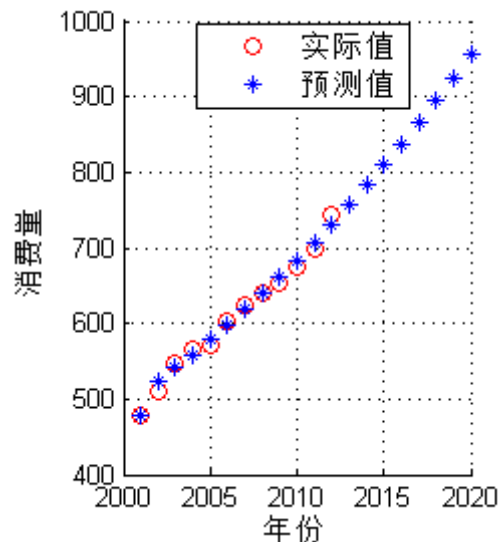


图 21 菜花消费量的实际值与预测值

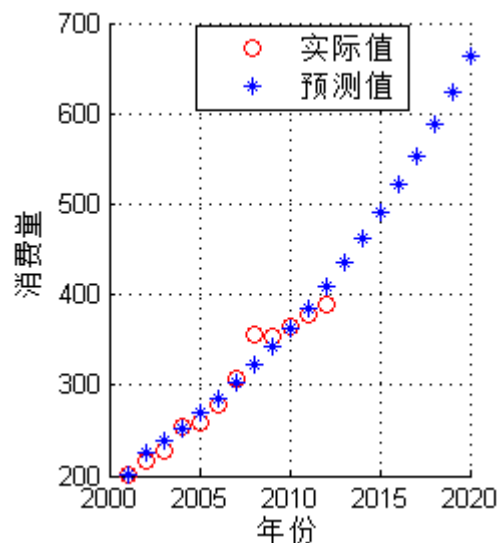


图 22 蘑菇消费量的实际值与预测值

由图 14—22 可以看出，根据灰色预测结果，到 2020 年消费量预测值一直处于增长趋势的蔬菜有土豆、西红柿、茄子、菠菜、胡萝卜、青椒、菜花和蘑菇，且西红柿、茄子、菠菜、胡萝卜和蘑菇基本处于曲线增长，其它处于直线增长；处于递减趋势的蔬菜只有大白菜，且近似为直线下降。

利用灰色预测模型，得到不同水果的预测结果如图 23—32 所示：

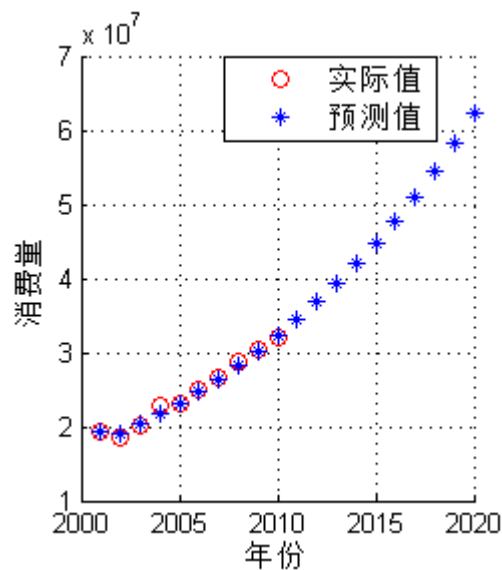


图 23 苹果消费量的实际值与预测值

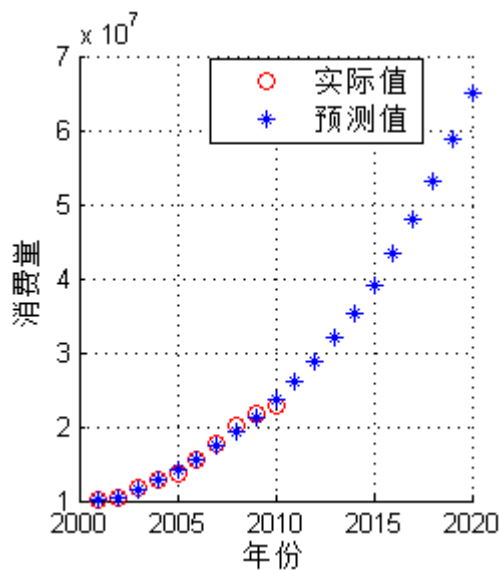


图 24 柑橘类消费量的实际值与预测值



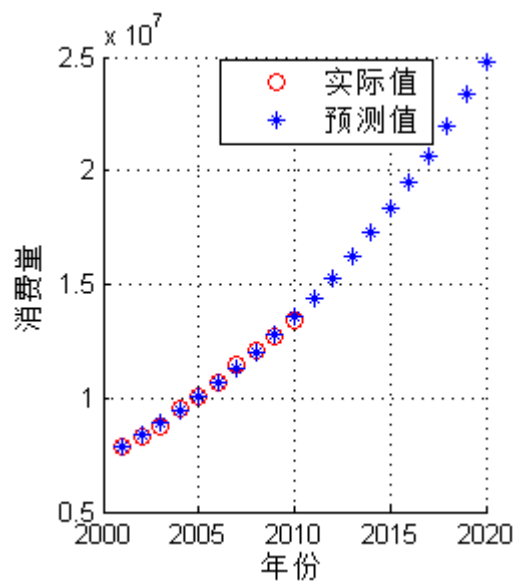


图 25 梨消费量的实际值与预测值

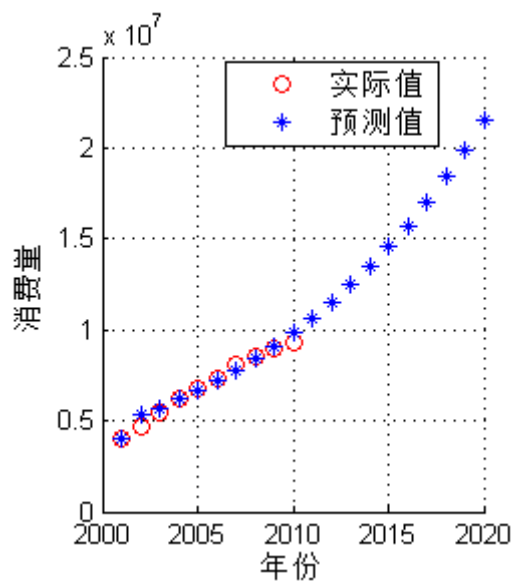


图 26 桃消费量的实际值与预测值

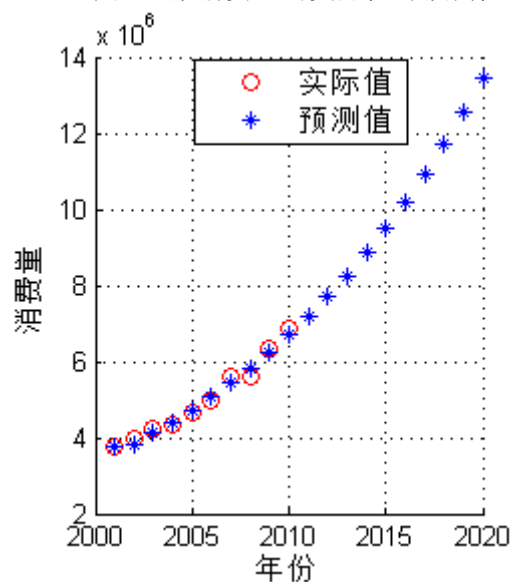


图 27 香蕉消费量的实际值与预测值

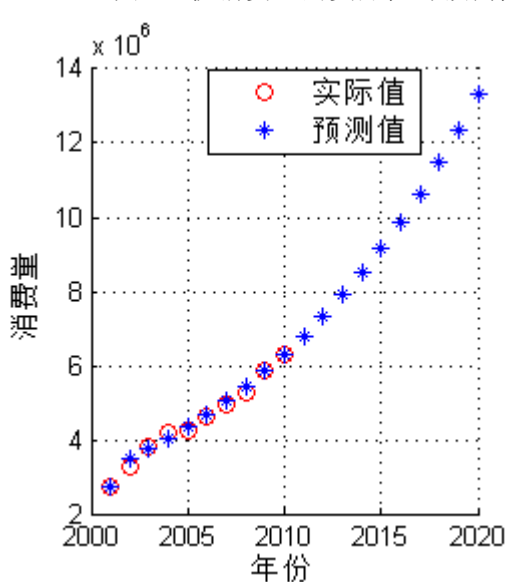


图 28 葡萄消费量的实际值与预测值

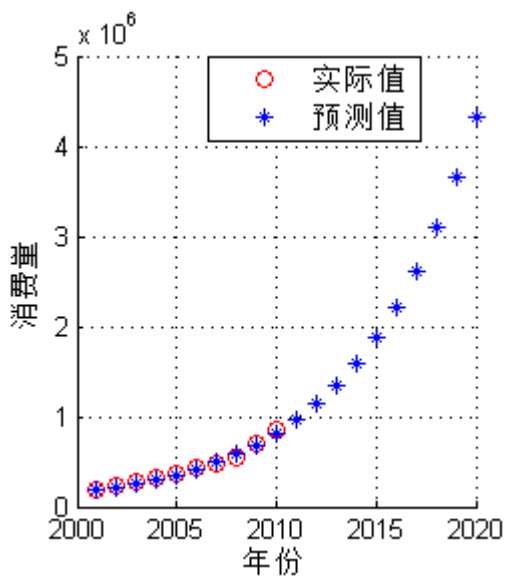
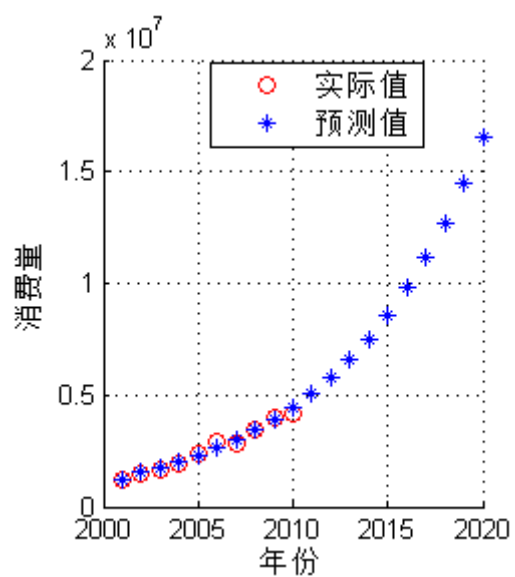


图 29 红枣消费量的实际值与预测值

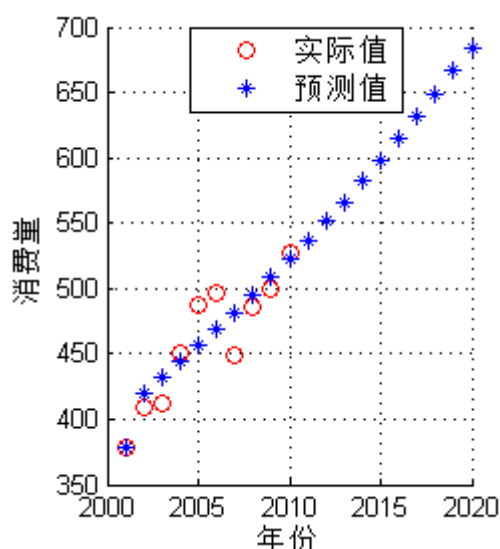


图 30 猕猴桃消费量的实际值与预测值

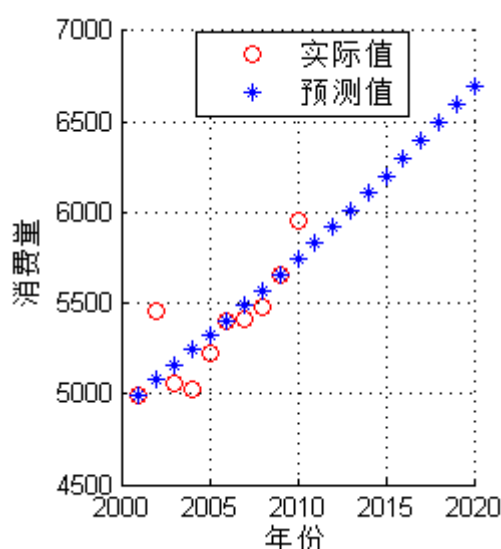


图 31 李子消费量的实际值与预测值

图 32 西瓜消费量的实际值与预测值

由图 23—32 可以看出，根据灰色预测结果，到 2020 年水果类消费量预测值均处于递增趋势，其中李子和西瓜的消费量近似为直线增长，其它水果均处于曲线增长。

## (2) BP 神经网络预测模型<sup>[12-15]</sup>的建立与求解

若记  $X_i = (x_1, x_2, \dots, x_i)$  是对某一研究对象一段时间的观测值，则对该时间序

列可以用下述模型来描述： $X_i = f(x_{i-1}, x_{i-2}, \dots, x_{i-n_0})$ ，其中  $f$  表示参数拟合函数，

$n_0$  为该模型的阶次。

任何时间序列都可以看作是一个由非线性机制确定的输入输出系统，时间序列的建模即是一个非线性参数拟合过程。1989 年 Robert Hocht-Nielson 证明了对于在闭区间内的任何一个连续函数都可以用含一个隐层的 BP 网络来近似，因而采用神经网络方法来实现非线性时间序列的建模与预测在理论上是可行的。典型的 BP 网络为三层前馈网络，包括输入层、隐层与输出层。

利用 BP 神经网络进行时间序列建模与预测的基本步骤主要分为三步：

### 1) 确定输入层的维数：

首先把时间序列分成两部分，第一部分大致为另一部分的两倍大小、开始窗口的大小可以任意选择、即输入神经元的数目可任定一初始值，用前一部分来训练网络、所得的网络用来预测第二部分，计算出预测误差。从小到大改变窗口，直至随着窗口尺寸的递增、预测精度不再有显著提高为止，此时的窗口尺寸即为输入层的维数。

### 2) 训练网络：

将全部序列作为训练样本，对该网络进行训练，得到该时间序列的神经网络预测模型。

### 3) 预测：利用所得到的模型进行预测。

运用 BP 神经网络进行时间序列建模与预测在实际应用中面临的主要问题，首先是 BP 网络的学习效率问题。由于 BP 算法采用梯度下降法来调整连接

权值，不可避免的会出现网络学习速度较慢，容易陷入局部极小或进入平坦区，从而导致无法收敛；同时，BP 网络就其本质而言，只是一非线性映射系统，而不是非线性动力系统，不具备动态适应能力，对于实时性要求高的系统难以满足；此外，为了得到满意的预测精度，对样本数据的准确性要求较高。

其次，从上述建模步骤可以看出，影响建模速度的因素主要有：1) 在定阶时，利用多组样本进行训练、预测、比较预测精度，进而得出网络输入层单元的数目，这势必要耗费大量的时间；2) 采用所有的样本进行训练，然后进行预测，样本数量的增加容易导致计算量的膨胀，特别是新增加一个样本时，必须将所有以前的样本加入重新训练，这不仅会使训练时间延长，而且，由于 BP 网络所容纳的信息是有限的，容易引起网络无法收敛。

针对这些问题，下面提出一种动态建模与预测方法，现以一步预测来介绍这种方法，多步预测方法可以类推。

BP 算法：

设  $x_i$  为一数据序列， $i=1,2,\dots,N$ ， $N$  为数据样本个数， $n_0$  为系统阶次，现在要用  $x_{i-1}, x_{i-2} \dots x_{i-n_0}$  来预测  $i$  时刻的值，故将  $x_{i-1}, x_{i-2} \dots x_{i-n_0}$  作为神经网络的输入，而  $x_i$  作为输出，利用 BP 算法对数据进行训练。BP 网络的输入神经元个数为  $n_0$ ，中间隐层神经元的个数为  $m$ ， $u_j$  为隐层神经元的输出， $w_{ij}$ ， $v_j$  分别为输入层与隐层、隐层与输出层之间的连接权值，数据样本数为  $N$ （下述公式中的上标  $n$  均表示第  $n$  个样本的相应数值）。

1) 各层神经元之间的输入输出关系满足：

$$u_j = f\left(\sum_{i=0}^{n_0} x_i w_{ij}\right), x_0 = -1 \quad (28)$$

$$y^n = \frac{1}{1 + \exp\left(\sum_{j=0}^m u_j v_j / \lambda^n\right)}, u_0 = 0 \quad (29)$$

$$f(s) = \frac{1}{1 + e^{-s}} - \frac{1}{2} \quad (30)$$

其中，若进入局部极小，则使  $\lambda^n > 1$ ，迅速脱离平坦区，在离开平坦区之后，再恢复  $\lambda^n = 1$ 。

2) 误差函数修正为：

$$E = \frac{1}{2} \left[ \sum_{n=1}^N (t^n - y^n)^2 + (1 - y^n / t^n)^2 \right] \quad (31)$$

其中  $t^n$  为第  $n$  个样本的实际值， $y^n$  为第  $n$  个样本由神经网络得出的预测值。

3) 权值的调节公式为：

$$\begin{cases} \delta_2^n = (y^n - t^n)(1/4 - (y^n)^2 + 0.05) \\ v_j(k+1) = v_j(k) - \eta \sum_{n=1}^N \delta_2^n u_j^n + \alpha \Delta v_j(k) \end{cases} \quad (32)$$

$$\begin{cases} \delta_{1j}^n = \delta_2^n v_j^n (1/4 - (u_j^n)^2) \\ w_{ij}(k+1) = w_{ij}(k) - \eta \sum_{n=1}^N \delta_{1j}^n x_i^n + \alpha \Delta w_{ij}(k) \end{cases} \quad (33)$$

$$\eta = \frac{\varphi}{N} \sum_{n=1}^N \left( \gamma / \sum_{j=0}^m u_j^n(k)^2 \right), \varphi = \beta \cdot \alpha \cdot p / c \quad (34)$$

当  $\Delta E > 0$  时，动量因子  $\alpha = 0$ ，而当  $\Delta E \leq 0$  时，取  $0 < \alpha < 1$ ； $\alpha$  为一接近 1 的常数， $c$  为一较大常数，通常取为 1500—4000， $p$  为收敛次数。

$$\beta = \begin{cases} 1 & \Delta E \leq 0 \\ \text{小于1的常数} & \Delta E > 0 \end{cases} \quad \Delta E = E(k+1) - E(k) \quad (35)$$

预测步骤：

为了提高预测的速度，本文将使用非线性时间序列的动态预测模型。所谓动态预测模型是指在建模过程中，BP网络的隐层单元数目、样本数目及输入层神经元数目（即窗口尺寸）不断变化，结构呈现出动态。在进行预测时，先选定一初始窗口尺寸、隐层神经元数目及样本数目，顺序移动窗口，进行训练与预测，同时不断检查预测效果，当预测精度不符合要求时，再动态地调整网络的结构与学习样本的数目。对于不断有新数据加入的时间序列，这种建模方法具有明显的优势。算法流程如图27。

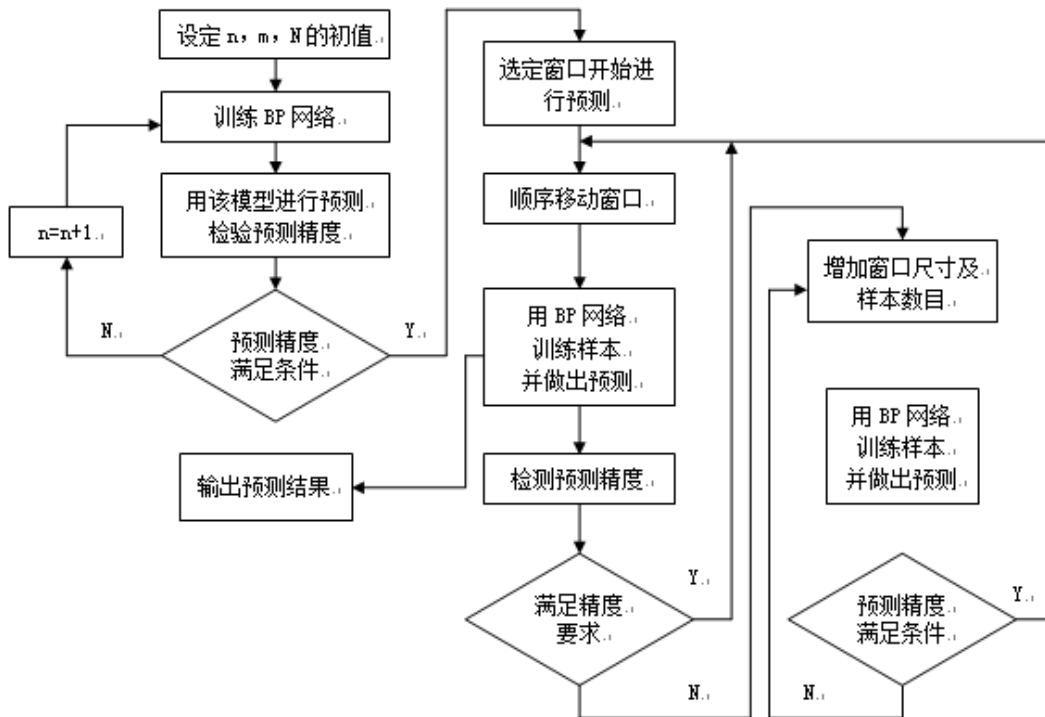


图33 BP神经网络算法流程图

利用 BP 神经网络预测模型，预测主要蔬菜和水果消费量，结果如图 34—52 所示：

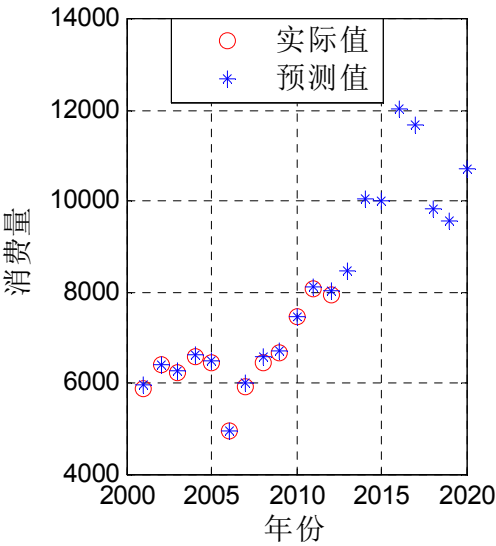


图 34 土豆消费量的实际值与预测值

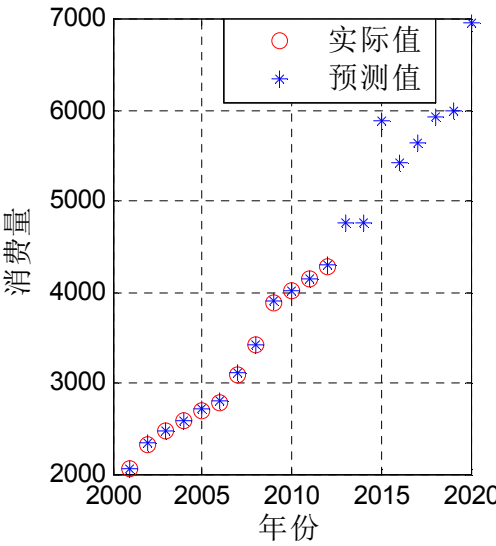


图 35 西红柿消费量的实际值与预测值

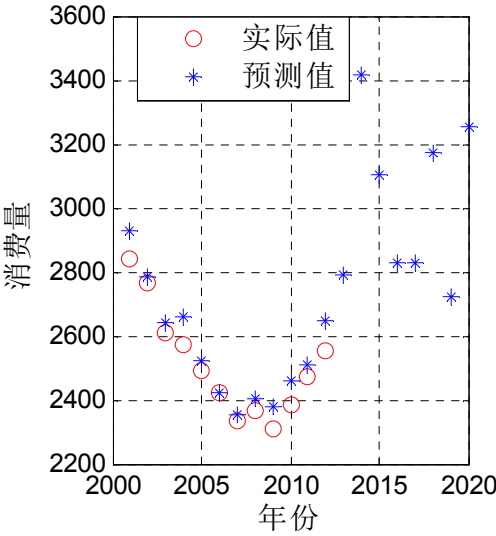


图 36 大白菜消费量的实际值与预测值

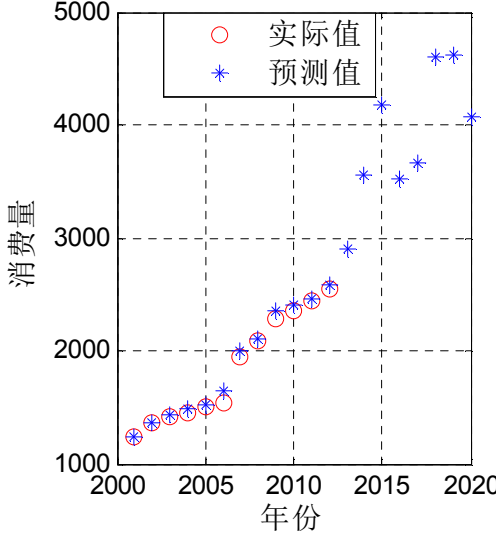


图 37 茄子消费量的实际值与预测值

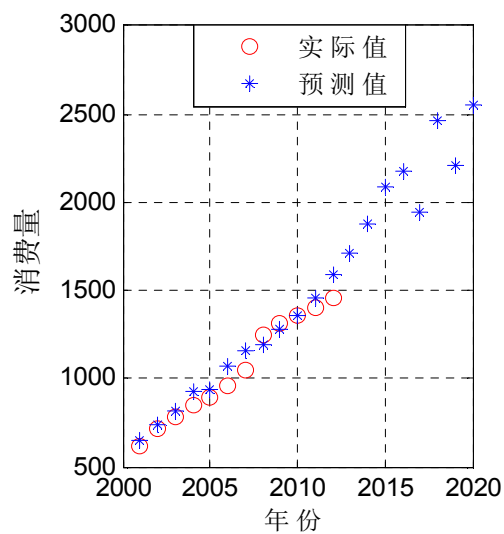


图 38 菠菜消费量的实际值与预测值

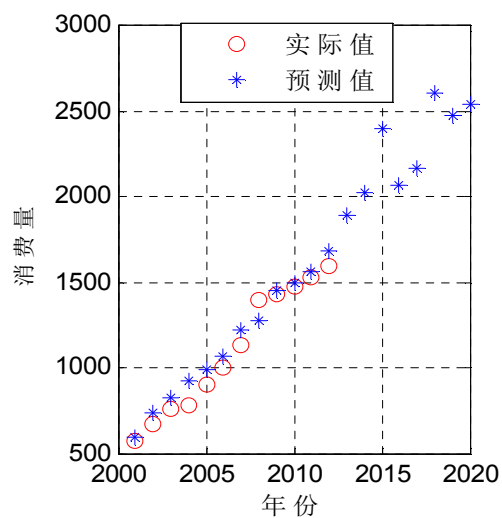


图 39 胡萝卜消费量的实际值与预测值

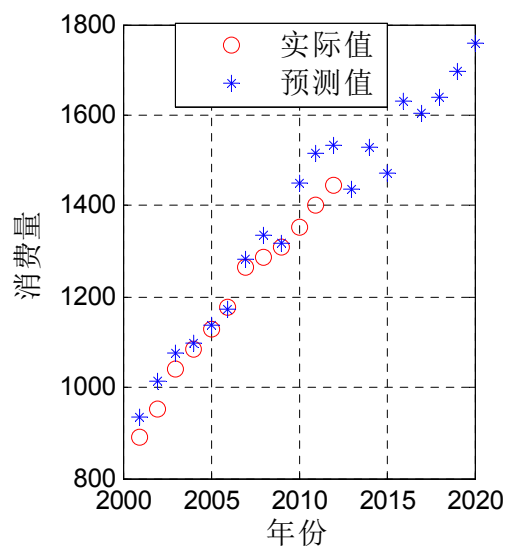


图 40 青椒消费量的实际值与预测值

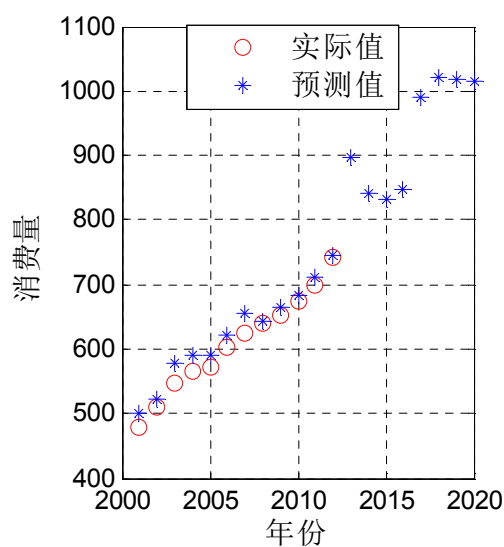


图 41 菜花消费量的实际值与预测值

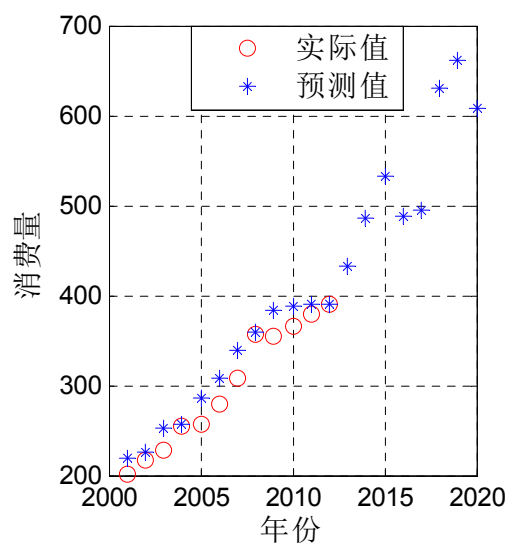


图 42 蘑菇消费量的实际值与预测值

由图 34—图 42 可以看出，根据 BP 神经网络预测，主要蔬菜的消费量除大白菜外，均呈现稳定的递增趋势，虽然有些蔬菜后来有些许波动，但是波动均不大，而大白菜呈现先增后减的趋势，并且在 2020 年前几年呈现不稳定性。

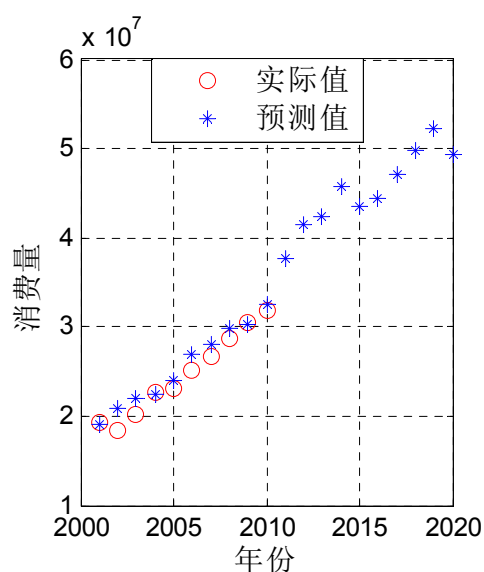


图 43 苹果消费量的实际值与预测值

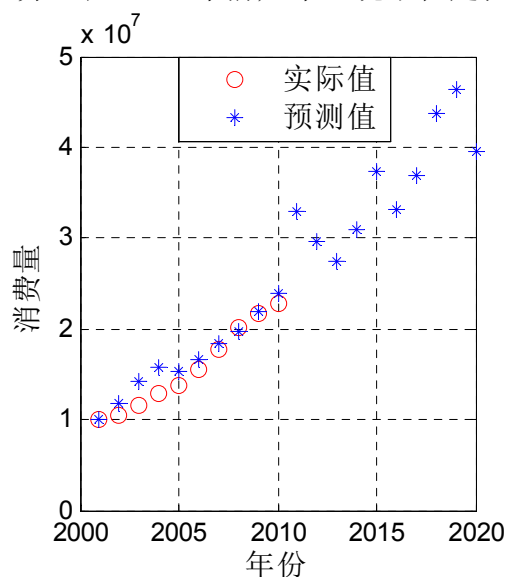


图 44 柑橘类消费量的实际值与预测值

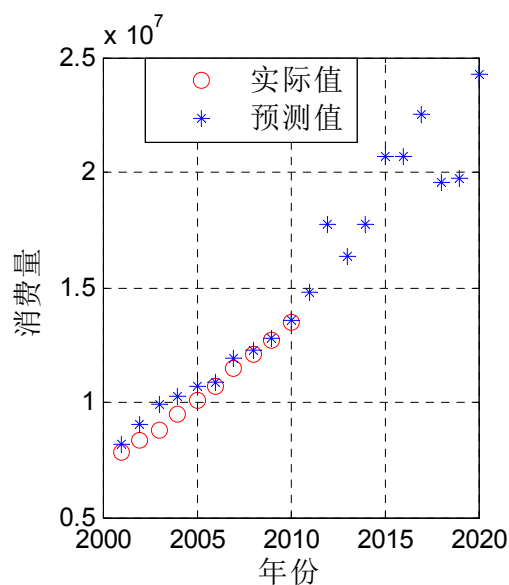


图 45 梨消费量的实际值与预测值

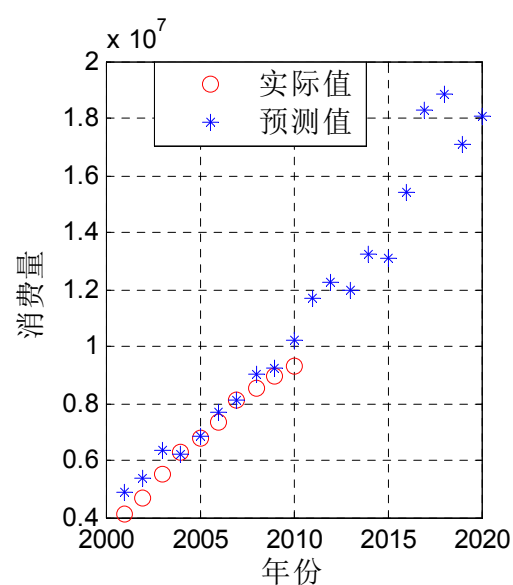


图 46 桃消费量的实际值与预测值

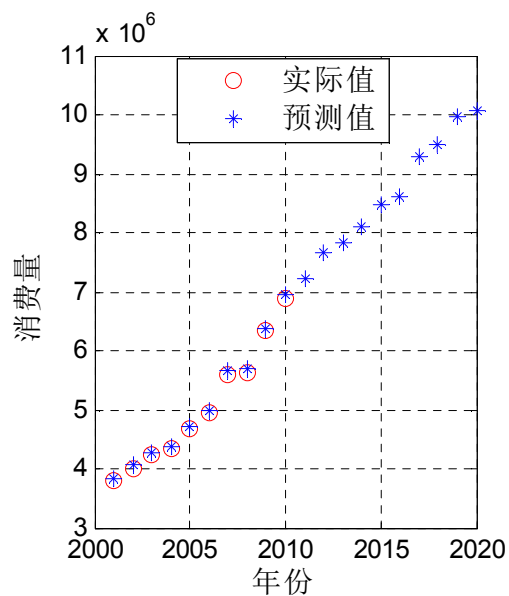


图 47 香蕉消费量的实际值与预测值

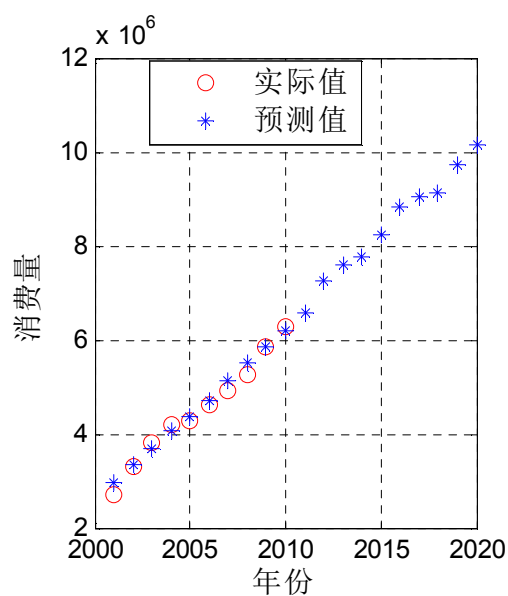


图 48 葡萄消费量的实际值与预测值

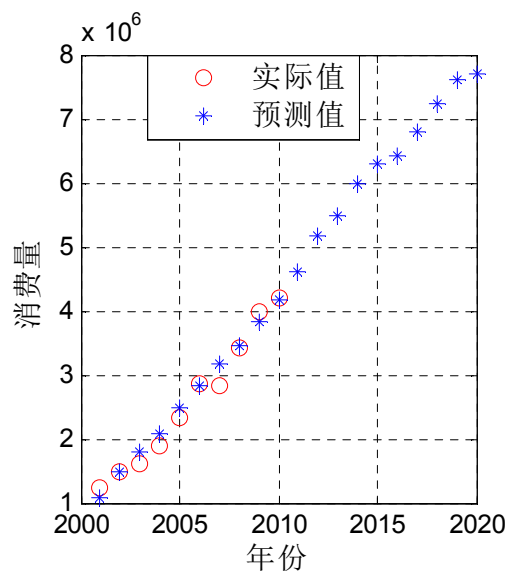


图 49 红枣消费量的实际值与预测值

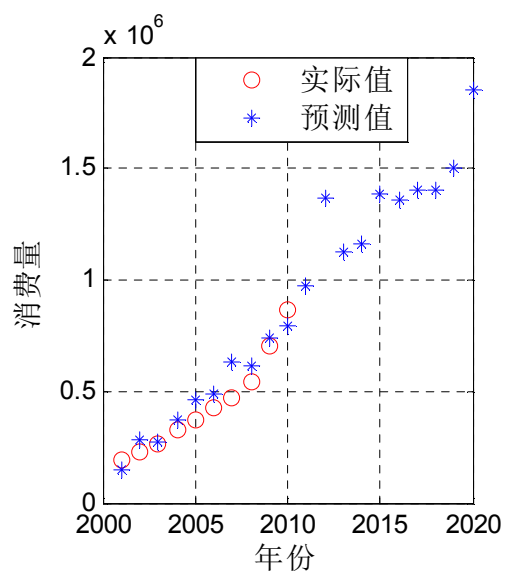


图 50 猕猴桃消费量的实际值与预测值



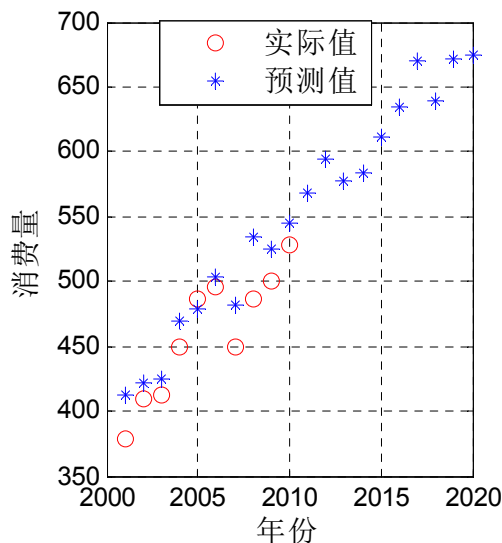


图 51 李子消费量的实际值与预测值

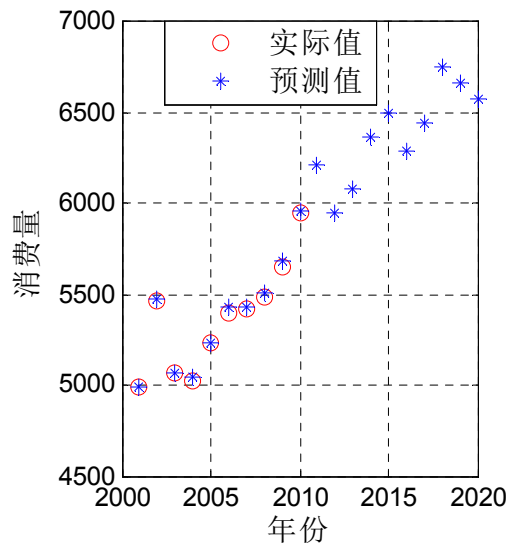


图 52 西瓜消费量的实际值与预测值

由图 43—图 52 可以看出，根据 BP 神经网络预测方法，主要水果的消费量预测中，所有水果均呈现上升趋势，其中苹果、香蕉、葡萄、红枣、猕猴桃和李子的上升趋势非常稳定，其它水果上升趋势有些许不稳定，但总体趋势一致。

### (3) 支持向量机回归预测模型<sup>[17-18]</sup>的建立与求解

回归的基本问题是发现一个函数  $f \in F$  ( $F$ —函数集)，使下面的期望风险函数最小，即： $R[f] = \int l(y - f(x)) dP(x, y)$ 。其中  $l(\cdot)$  为损失函数，表示  $y$  和  $f(x)$  之间的偏差，其常用形式为  $l(\cdot) = |y - f(x)|^p$ ，其中  $p$  为某个正整数。因为  $P(x, y)$  不能事先知道，所以不能利用上式直接计算  $R[f]$ 。

根据结构风险最小化，有  $R[f] \leq R_{emp} + R_{gen}$  式中  $R_{emp} = \frac{1}{l} \sum_{i=1}^l l(y_i - f(x_i))$  为经验风险； $R_{gen}$  为  $f(x)$  复杂度的一种度量。因此，可用  $R_{emp} + R_{gen}$  来确定  $R[f]$  的上限。

SVM 用来解决回归预测的基本思想如下：给定以  $P(x, y)$  为概率的观测样本集  $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \in R^n \times R$ 。设回归函数为：

$F = \{f | f(x) = w^T \Phi(x) + b, w \in R^n\}$ 。引进下述结构风险函数：

$$R_{reg} = \frac{1}{2} \|w\|^2 + C \cdot R_{emp}^{\epsilon}[f] \quad (36)$$

式中  $\|w\|^2$  为描述函数  $f(\cdot)$  为复杂度的项； $C$  为常数。它的作用是在经验风

险和模型复杂度之间取一折中。引进不灵敏损失函数  $\varepsilon$ ，具体定义

$$|y - f(x)|_{\varepsilon} = \begin{cases} 0 & , |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{其他} \end{cases} \quad (37)$$

它意味着不惩罚偏差小于  $\varepsilon$  的项，这样可以增加回归的鲁棒性。

$R_{emp} = \frac{1}{l} \sum_{i=1}^l |y_i - f(x_i)|_{\varepsilon}$  对式 (14) 最小化，体现了统计学习理论的核心思想，

既控制训练误差，又控制模型复杂度，目的是为了获得一个小的期望风险。这样可以提高模型的泛化能力。

上述函数回归问题等价于最小化代价泛函

$$\begin{aligned} \min & \frac{1}{2} w^T w + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \\ \text{s.t.} & \quad y_i - w^T x_i - b \leq \varepsilon + \zeta_i \\ & \quad w^T x_i + b - y_i \leq \varepsilon + \zeta_i^* \\ & \quad \zeta_i, \zeta_i^* \geq 0 \end{aligned} \quad (38)$$

式中  $\zeta_i, \zeta_i^*$  为引进的松弛变量，目的是使式 (16) 的解存在。

利用拉格朗日函数和对偶原理，可以得到式 (17) 的对偶问题

$$\begin{aligned} \min_{\alpha, \alpha^*} & \left\{ \frac{1}{2} [\alpha, (\alpha^*)^T] \begin{bmatrix} Q & -Q \\ -Q & Q \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} + [\varepsilon I^T + y^T \quad \varepsilon I^T - y^T] \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} \right\} \\ \text{s.t.} & \quad [I^T, -I^T] \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (39)$$

式中  $Q_{i,j} = \Phi^T(x_i) \Phi(x_j)$ ;  $I = [1, \dots, 1]^T$ ;  $\alpha, \alpha^*$  为 Langrange 乘子。

求解此二次型规划可求得  $\alpha$  的值，同时求得  $w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i)$  (40)

利用 KKT (Karush-Kuhn-Tucker) 条件可以计算出常值偏差  $b$ ,

$$\begin{cases} b = y_j - \varepsilon - \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_j, x_i), \alpha_i, \alpha_i^* \in [0, C] \\ b = y_j + \varepsilon - \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_j, x_i), \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (41)$$

根据上述推导可得回归函数  $f(x)$  的表达式为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_j, x_i) + b \quad (42)$$

其中， $K(x_j, x_i) = \Phi^T(x) \Phi(x_i)$  为一个满足 Mercer 条件的核函数。该函数可

在不知非线性变换的具体形式下实现算法的非线性化，这是支持向量机的一个显著特点。

常用的核函数为：

阶次为 $b$ 的多项式核函数  $K(x_i, x_j) = (x_i \cdot x_j + 1)^d$

径向基函数核函数  $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma^2)^d$

神经网络核函数  $K(x_i, x_j) = \tanh[c_1(x_i \cdot x_j) + c_2]^d$

利用支持向量机回归模型，预测主要蔬菜和水果消费量如图 53—71 所示：

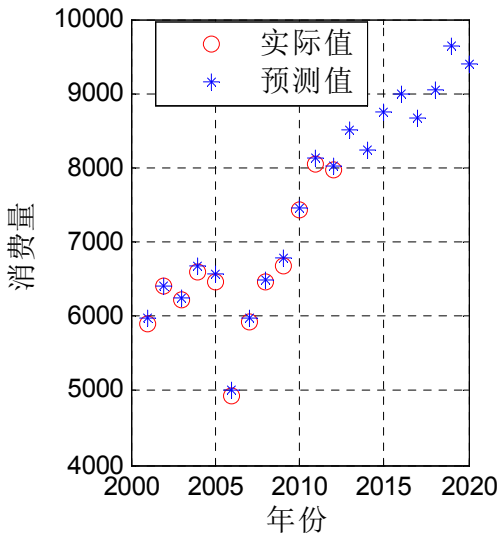


图 53 土豆消费量的实际值与预测值

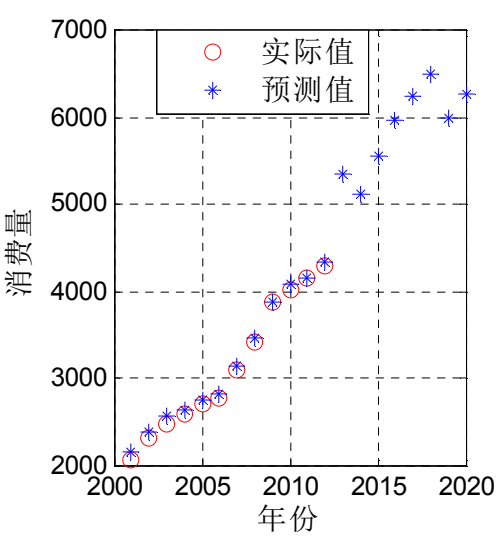


图 54 西红柿消费量的实际值与预测值

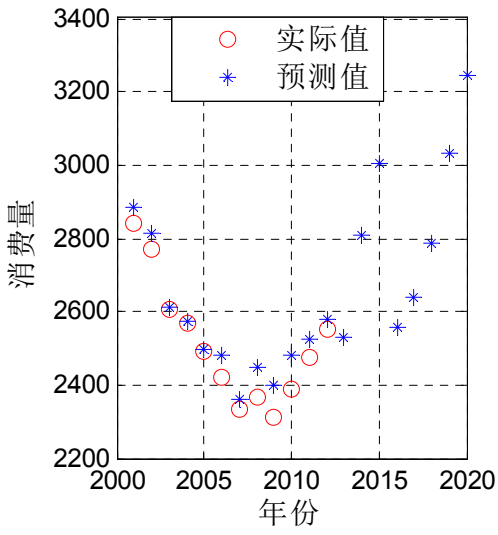


图 55 大白菜消费量的实际值与预测值

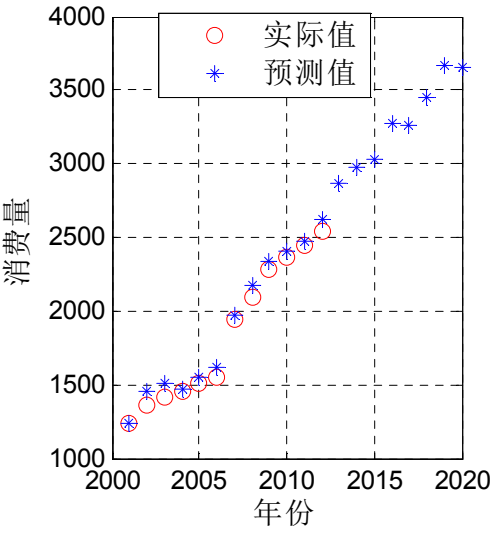


图 56 茄子消费量的实际值与预测值

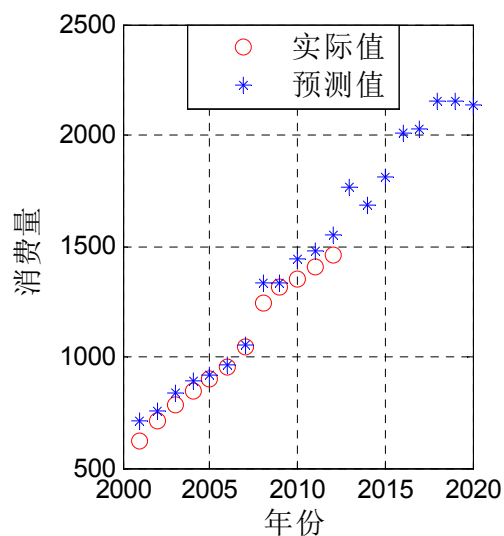


图 57 菠菜消费量的实际值与预测值

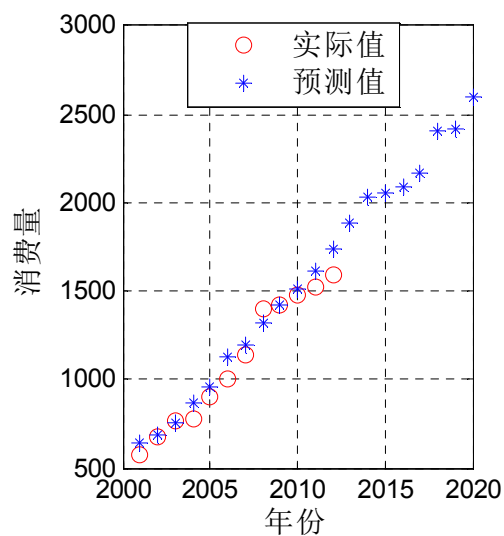


图 58 胡萝卜消费量的实际值与预测值

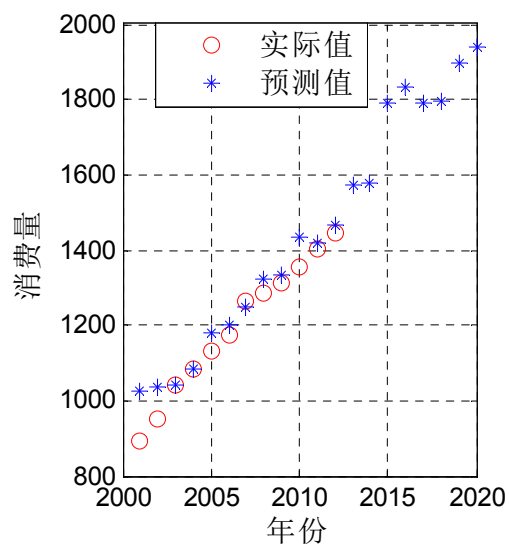


图 59 青椒消费量的实际值与预测值

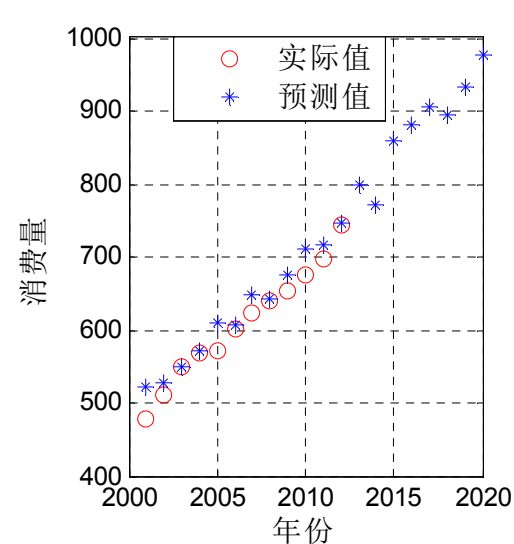


图 60 菜花消费量的实际值与预测值

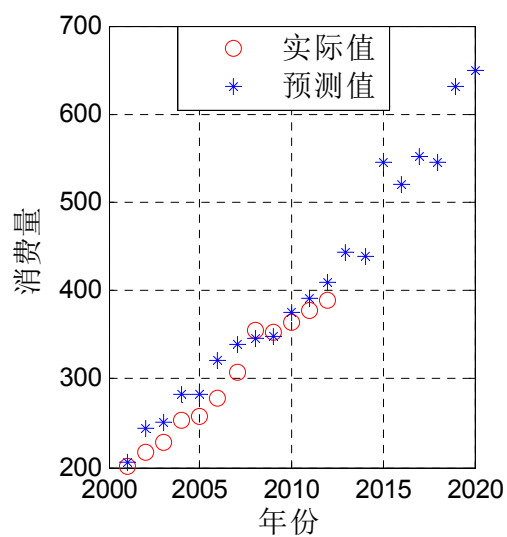


图 61 蘑菇消费量的实际值与预测值

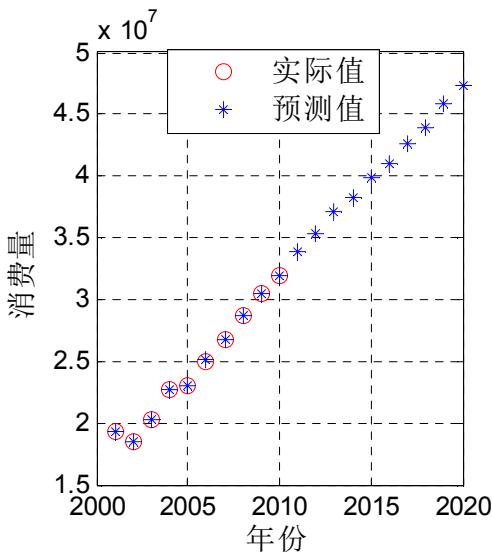


图 62 苹果消费量的实际值与预测值

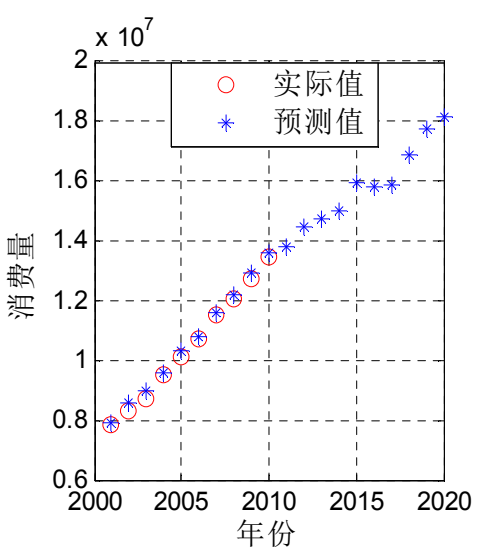


图 64 梨消费量的实际值与预测值

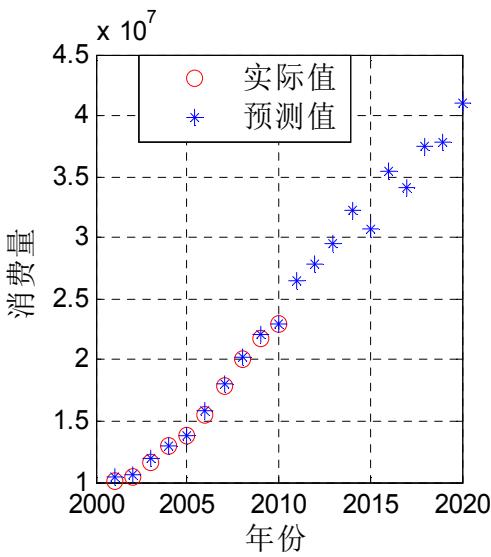
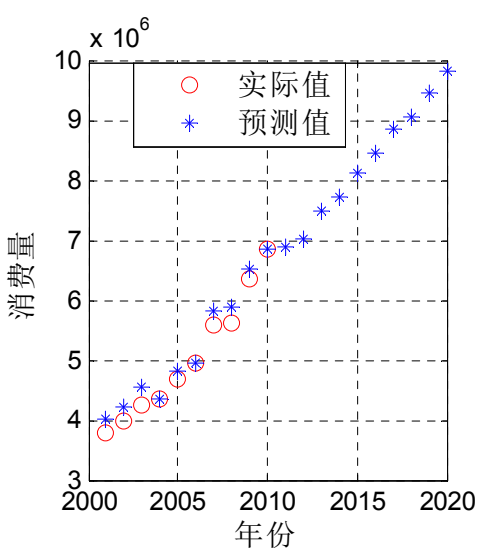


图 63 柑橘类消费量的实际值与预测值

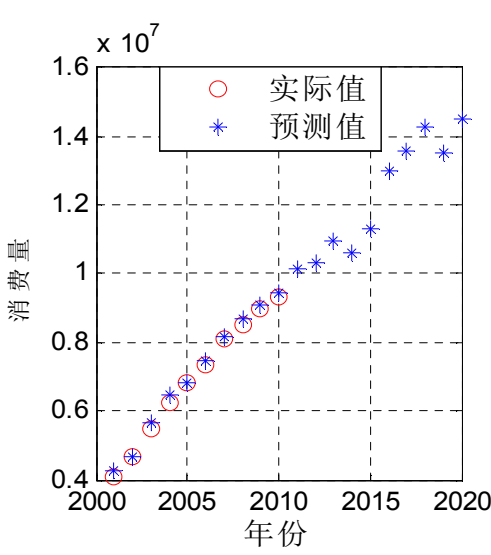


图 65 桃消费量的实际值与预测值

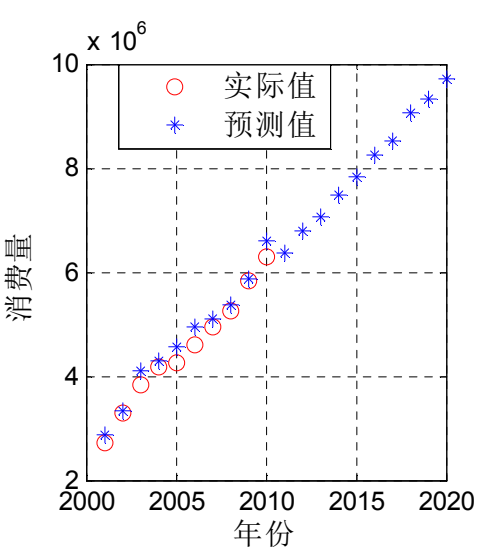


图 66 香蕉消费量的实际值与预测值

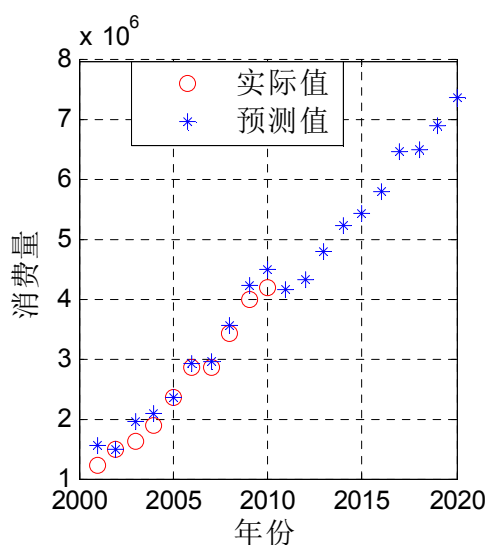


图 67 葡萄消费量的实际值与预测值

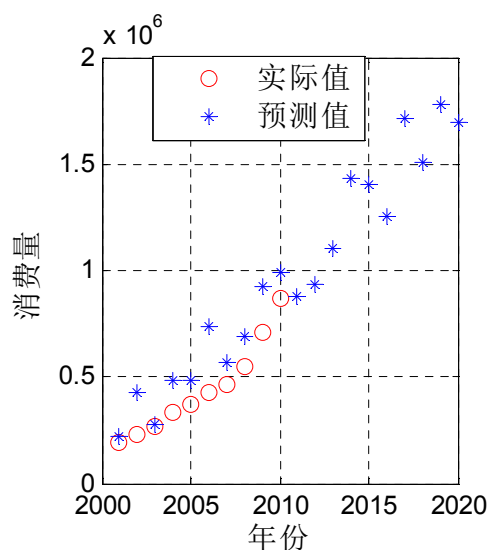


图 68 红枣消费量的实际值与预测值

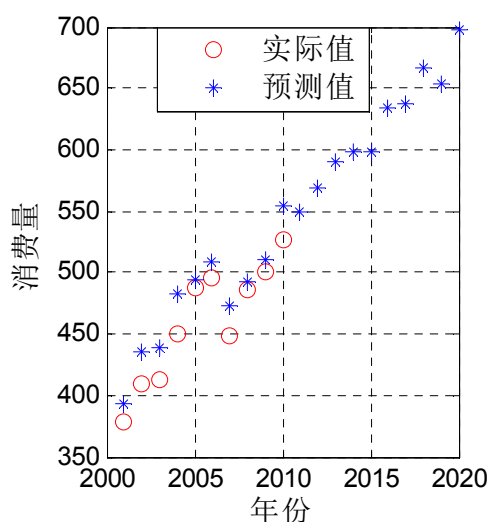


图 69 猕猴桃消费量的实际值与预测值

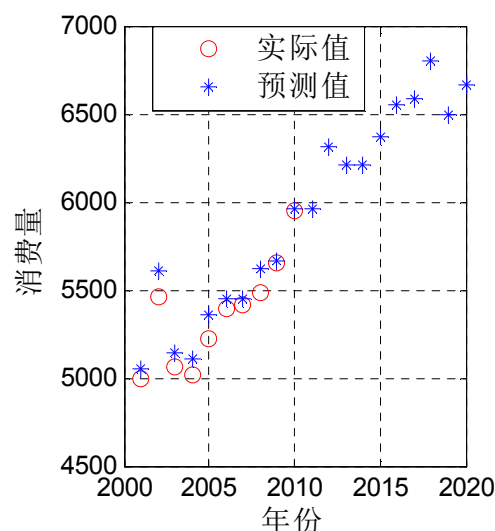


图 70 李子消费量的实际值与预测值

图 71 西瓜消费量的实际值与预测值

由图 53—71 可以看出，根据支持向量机回归模型，主要蔬菜和水果的变化趋势大致与灰色预测模型和 BP 神经网络预测模型类似，蔬菜中除大白菜外，其它均呈现比较稳定的上升趋势，而大白菜呈现先减后增的趋势，水果均呈现比较稳定的上升趋势。

#### (4) 基于 WPA 算子的组合预测模型

由于需要对 19 种果蔬进行趋势预测，单项预测方法无法同时对 19 种果蔬进行准确预测。所以这里采用基于加权 power 平均 (WPA) 算子的组合预测模型<sup>[19-22]</sup>，能够有效提高对各种果蔬预测准确率。

**定义 2.** 设  $WPA: R^n \rightarrow R$  为  $n$  元函数， $W = (w_1, w_2, \dots, w_n)^T$  是与 WPA 有关的加权向量，满足  $\sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, 2, \dots, n$ ，若：

$$WPA(a_1, a_2, \dots, a_n) = \frac{\sum_{i=1}^n w_i(1+T(a_i))a_i}{\sum_{i=1}^n w_i(1+T(a_i))} = \sum_{i=1}^n \frac{w_i(1+T(a_i))}{\sum_{i=1}^n w_i(1+T(a_i))} a_i \quad (43)$$

则称函数 WPA 是  $n$  维加权 Power 算子，简称 WPA 算子，其中：

$$T(a_i) = \sum_{\substack{j=1 \\ j \neq i}}^n w_j Sup(a_i, a_j) \quad (44)$$

$Sup(a_i, a_j)$  表示  $a_j$  对  $a_i$  的支撑度，并且满足以下三个性质：

- (1)  $Sup(a_i, a_j) \in [0, 1]$ ;
- (2)  $Sup(a_i, a_j) = Sup(a_j, a_i)$ ;
- (3) 如果  $|a_i - a_j| < |a_s - a_t|$ ，则有  $Sup(a_i, a_j) \geq Sup(a_s, a_t)$ 。

若记  $\tilde{w}_i = \frac{w_i(1+T(a_i))}{\sum_{i=1}^n w_i(1+T(a_i))}$ ，则式 (21) 可写成：

$$WPA(a_1, a_2, \dots, a_n) = \sum_{i=1}^n \tilde{w}_i a_i \quad (45)$$

其中  $\tilde{w}_i$  可以看作是一种复合权重，它不仅考虑到数据本身的重要性，还考虑到了数据之间的关系。

特别地，当  $w = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})^T$ ，这时 WPA 算子就退化成 PA 平均算子<sup>[27]</sup>。

### 基于 WPA 算子的组合预测模型

设对同一预测对象的某个指标序列为  $\{x_t, t=1, 2, \dots, N\}$ ，存在  $m$  种单项预测方法对其进行预测，设  $x_{it}$  为第  $i$  种单项预测模型在第  $t$  时刻的预测值， $i=1, 2, \dots, m$ ， $t=1, 2, \dots, N$ 。

令

$$\hat{x}_t = WPA(x_{1t}, x_{2t}, \dots, x_{mt}) = \frac{\sum_{i=1}^m w_i(1+T(x_{it}))x_{it}}{\sum_{i=1}^m w_i(1+T(x_{it}))} \quad (46)$$

其中  $\hat{x}_t$  为第  $t$  时刻在 WPA 算子下的组合预测值， $w_i$  ( $i=1, 2, \dots, m$ ) 为第  $i$  种单项预

测方法的加权系数，且  $w_i$  满足

$$\begin{cases} \sum_{i=1}^m w_i = 1 \\ w_i \geq 0 \quad i = 1, 2, \dots, m \end{cases} \quad (47)$$

定义 3. 令

$$e_t = x_t - \hat{x}_t = \sum_{i=1}^m \frac{w_i(1+T(x_{it}))}{\sum_{i=1}^m w_i(1+T(x_{it}))} e_{it} = \sum_{i=1}^m \tilde{w}_i e_{it} \quad (48)$$

则称  $e_t$  为组合预测第  $t$  时刻在 WPA 算子下的预测误差， $e_{it} = (x_t - x_{it})$  为第  $i$  种单项预测方法第  $t$  时刻在 WPA 算子下的预测误差，其中  $t = 1, 2, \dots, N$ 。

设  $J$  表示为  $N$  期总得组合预测误差平方和，则有：

$$J = \sum_{t=1}^N e_t^2 = \sum_{t=1}^N (x_t - \hat{x}_t)^2 = \sum_{t=1}^N \sum_{i=1}^m \sum_{j=1}^m \tilde{w}_i \tilde{w}_j e_{it} e_{jt} \quad (49)$$

因此，可得到以组合预测误差平方和为准则的组合预测模型为下列最优化问题：

$$\begin{aligned} \min J = \min \sum_{t=1}^N e_t^2 &= \sum_{t=1}^N (x_t - \hat{x}_t)^2 = \sum_{t=1}^N \sum_{i=1}^m \sum_{j=1}^m \tilde{w}_i \tilde{w}_j e_{it} e_{jt} \\ s.t. \quad &\begin{cases} \sum_{i=1}^m \tilde{w}_i = 1 \\ \tilde{w}_i \geq 0 \quad i = 1, 2, \dots, m \end{cases} \end{aligned} \quad (50)$$

记  $W = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_m)^T$ ,  $R = (1, 1, \dots, 1)^T$ ,  $e_i = (e_{i1}, e_{i2}, \dots, e_{iN})^T$ ，则  $W$  表示组合预测加权系数列向量， $R$  表示元素全为 1 的  $m$  维列向量， $e_i$  表示第  $i$  种单项预测方

法的预测误差列向量。再令： $E_{ij} = e_i^T e_j = \sum_{t=1}^N e_{it} e_{jt}$ ,  $i, j = 1, 2, \dots, m$ ，则称  $E = (E_{ij})_{m \times m}$

为在 WPA 算子下  $m$  阶组合预测误差信息矩阵。因此式 (28) 可以写成如下矩阵形式：

$$\begin{aligned} \min J &= W^T E W \\ s.t. \quad &\begin{cases} R^T W = 1 \\ W \geq 0 \end{cases} \end{aligned} \quad (51)$$

利用该组合预测模型可得各种果蔬组合预测的权值，结果见表 9，利用表 9 的结果，考虑前方三种单项预测方法的预测结果，得到各种果蔬的组合销量趋势，结果如图 72-73 所示：



表 9 各种水果和蔬菜的组合预测权系数

	无偏灰色预测模型权值	BP 神经网络预测模型权值	支持向量回归预测模型权值
土豆	0.0000	0.8714	0.1276
西红柿	0.0354	0.9636	0.0000
大白菜	0.2124	0.5111	0.2755
茄子	0.1676	0.7746	0.0568
菠菜	0.6299	0.0000	0.3690
胡萝卜	0.2950	0.3011	0.4029
青椒	0.9832	0.0000	0.0158
菜花	0.8332	0.1658	0.0000
蘑菇	0.7376	0.2614	0.0000
苹果	0.008	0.000	0.991
柑橘类	0.249	0.000	0.750
梨	0.855	0.000	0.144
桃	0.117	0.000	0.882
香蕉	0.194	0.805	0.000
葡萄	0.842	0.044	0.112
红枣	0.442	0.375	0.182
猕猴桃	0.921	0.000	0.078
李子	0.569	0.000	0.430
西瓜	0.001	0.998	0.000

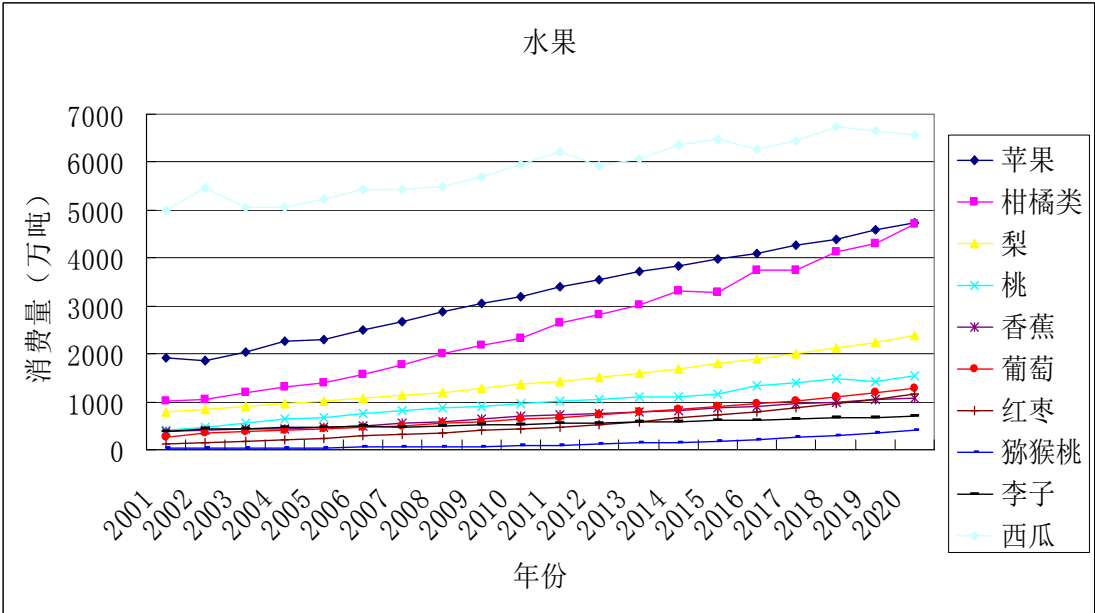


图 72 各种水果的消费量趋势

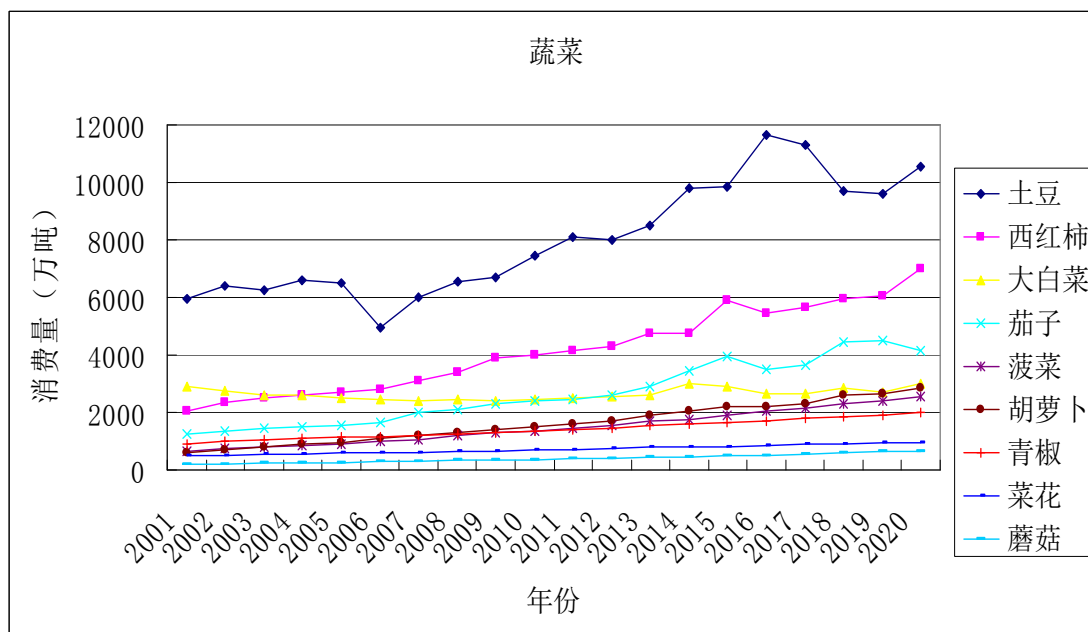


图 73 各种蔬菜的消费量趋势

由图 72 可以看出，根据组合预测模型，主要水果消费量均呈现上升趋势，其中苹果、柑橘类、梨增长趋势明显，增长幅度最明显的是柑橘类，而西瓜消费量会有一定的短期波动。由图 73 可以看出，根据组合预测模型，主要蔬菜的消费量除大白菜有一定的波动呈现下降趋势，其它蔬菜均呈现不同程度的增长，其中增长趋势最明显的是西红柿和土豆，菜花和蘑菇的增长趋势最弱。

## 4.2 问题二的模型建立与求解

### 4.2.1 中国居民营养摄入水平分析

在这里我们需要计算各种营养素的人均摄入水平，根据问题一已求出的各种果蔬的消费量，已知人口数，可以直接得到我国营养摄入现状，结合营养素的标准范围，评价是否合理。问题一又预测了至 2020 年果蔬的消费量，我们只需预测至 2020 年的人口数量，就可以得到人均消费量，与标准营养素区间进行比较，便可以得到我国人体营养状况的发展趋势。

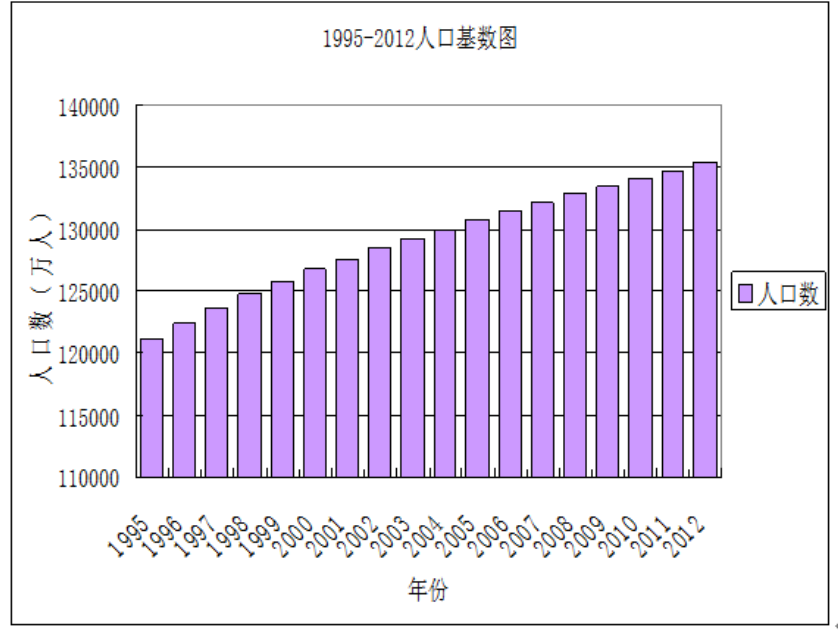


图 74 1995-2012 年人口基数图

根据图 74 中人口数据，应用问题一中的无偏灰色模型对人口进行预测得到结果如图 75 所示：

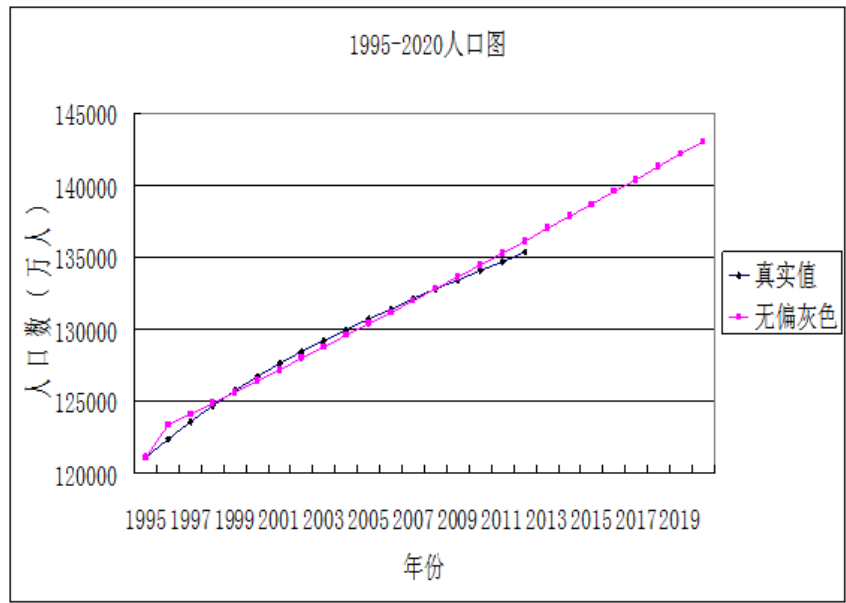


图 75 人口预测值与真实值

在问题一中，可以知道在  $t$  年第  $i$  种果蔬的年消耗量的实际值或未来预测值，记为  $C_{it}$ ，又人口在  $t$  年的实际值或未来预测值已知，设为  $P_t$ 。则在  $t$  年第  $i$  种果蔬的日人均消费量为：

$$c_{it} = \frac{C_{it}(1 - S_3)}{365P_t} \quad (52)$$

其中  $S_3$  为零售市场到餐桌的损耗率。

计算结果如下图所示：

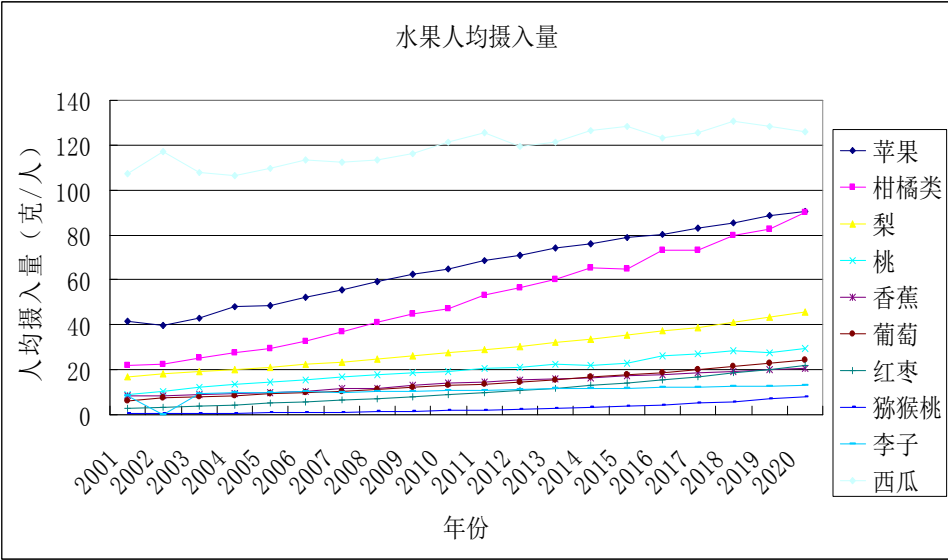


图 76 水果的人均摄入量

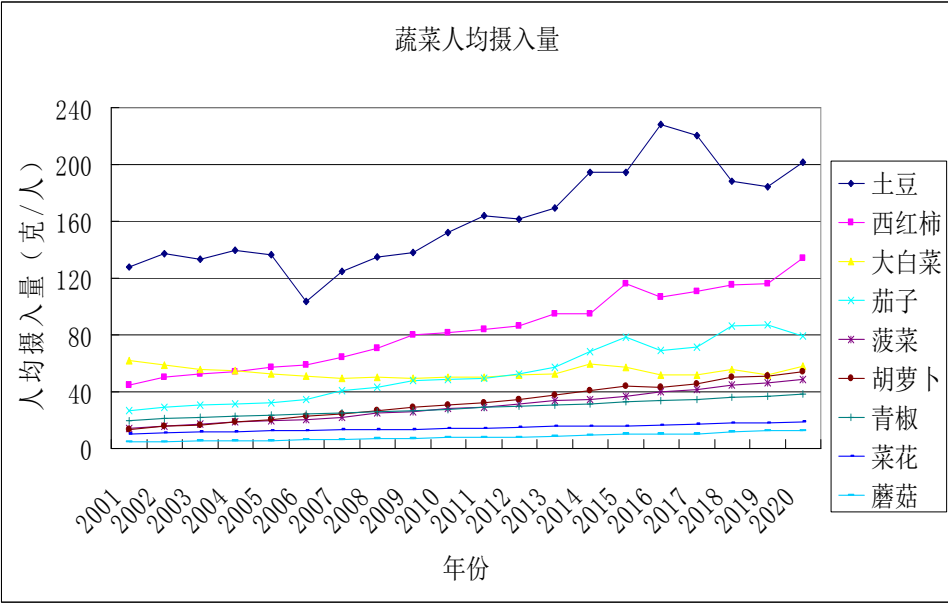


图 77 人口预测值与真实值

则在  $t$  年第  $j$  种营养素的人均每日摄入量为：

$$A_{jt} = \frac{\sum_i c_{it} a_{ij}}{\delta_j}, \quad (53)$$

式中  $a_{ij}$  表示第  $i$  种果蔬品种第  $j$  种营养素的含量； $\delta_j$  为参数，控制第  $j$  种营养素在果蔬中摄取量占总摄取量的比重，因为人体所需的营养素不仅来自果蔬，还有主食和肉类等。

计算结果如下图 72-83：

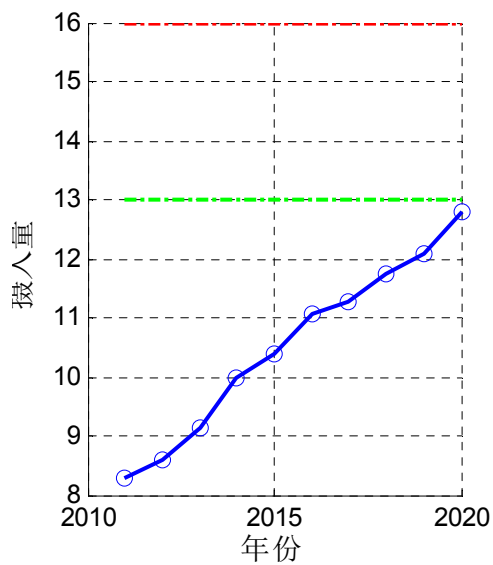


图 78 锌的摄入量与标准区间

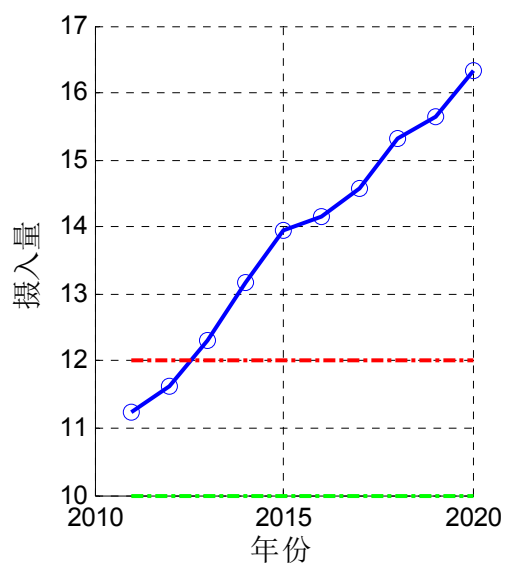
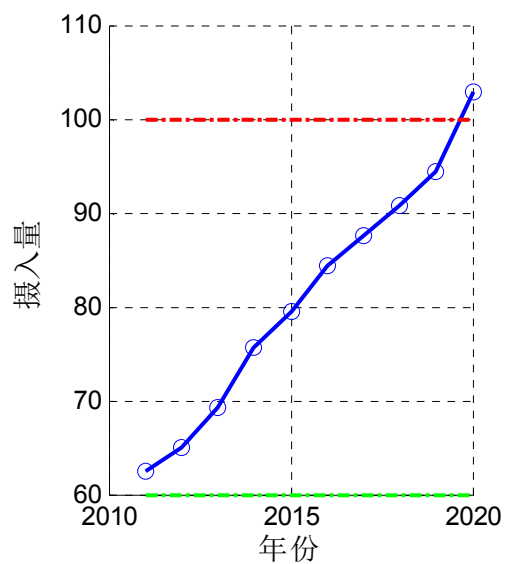


图 79 维生素 E 的摄入量与标准区间



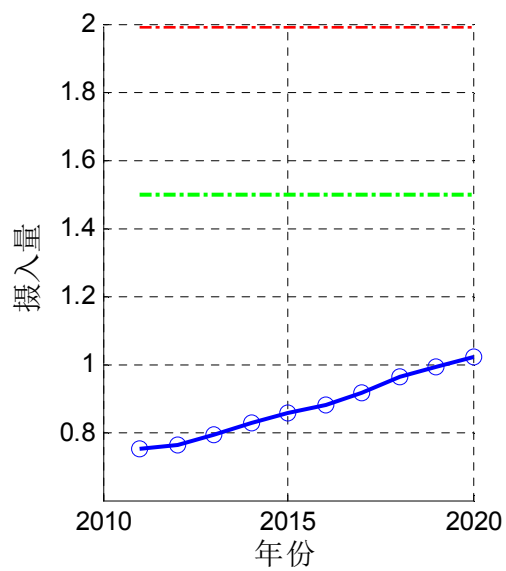
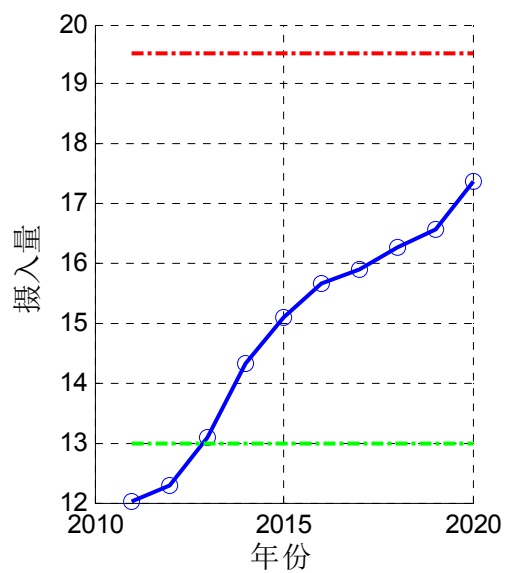


图 80 维生素 C 的摄入量与标准区间

图 81 维生素 B6 的摄入量与标准区间



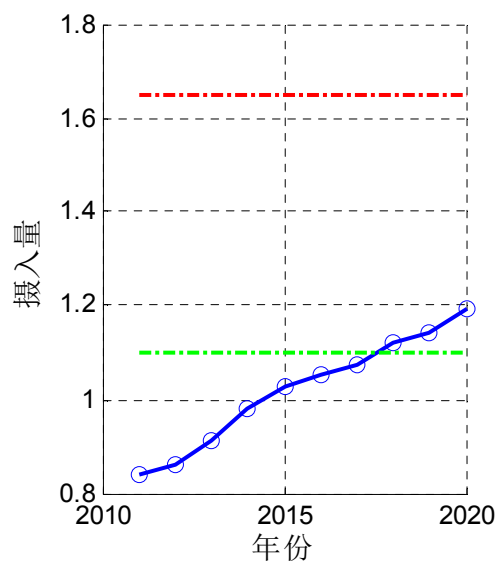
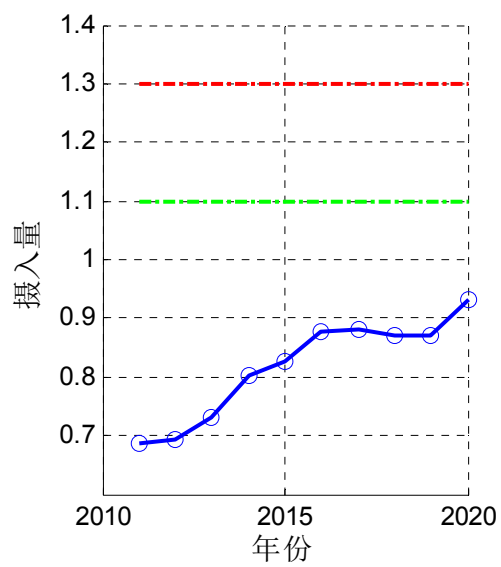


图 82 维生素 B3 的摄入量与标准区间

图 83 维生素 B2 的摄入量与标准区间



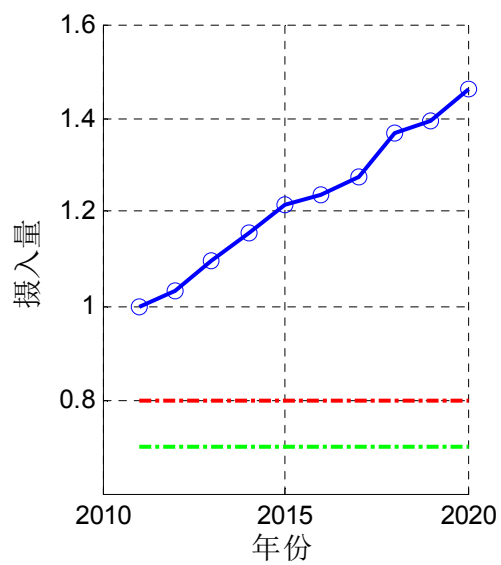


图 84 维生素 B1 的摄入量与标准区间



图 85 维生素 A 的摄入量与标准区间

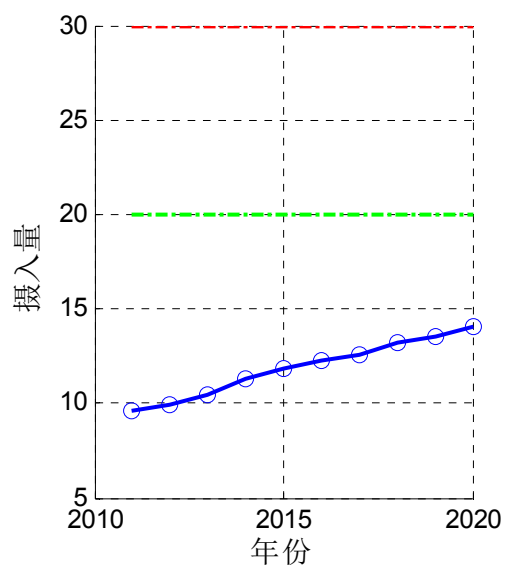


图 86 膳食纤维的摄入量与标准区间

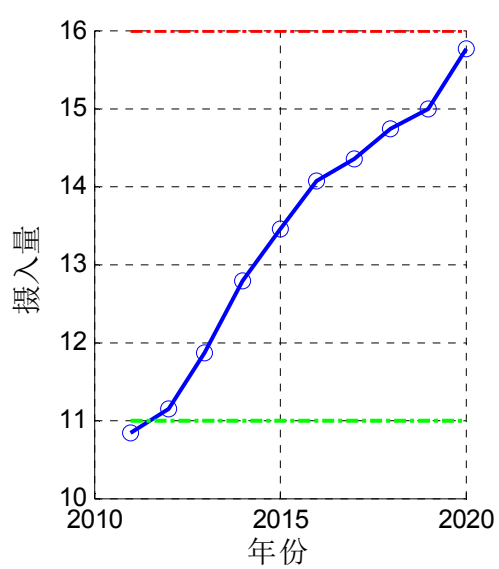


图 87 铁的摄入量与标准区间



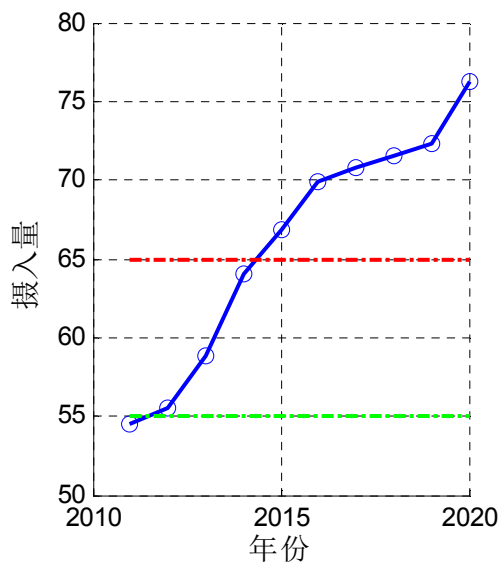


图 88 蛋白质的摄入量与标准区间

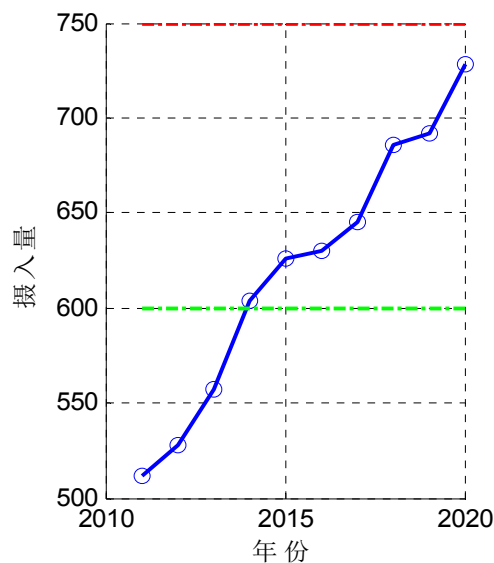


图 89 钙的摄入量与标准区间

由图 78—89，我们可以看出目前我国各种营养素的摄入现状为：锌、维生素 B6、维生素 B2、维生素 B1、膳食纤维这些营养素的摄入量比标准范围较小。维生素 E 和维生素 A 的这些营养素的摄入量高于标准范围，其他营养素的摄入量趋于正常。

从图中还可以看出，随着时间的增长，每种营养素的摄入量都在增加。这是因为我国作为发展中国家，虽然人民的生活水平在不断提高，但是大部分人还不具备营养均衡的观念与意识，因此在未来的很长时间内每种营养素的摄入量还会呈增长趋势。

另外，上述各图也可以定性反映出，到 2020 年，锌、维生素 B6、维生素 B2、维生素 B1、膳食纤维这些营养素的摄入量逐渐接近或达到标准营养区间，但是维生素 E 和维生素 A 的摄入量和现状一样继续高于标准范围，而现状维持正常摄入量的蛋白质将会在未来高于标准。

#### 4.2.2 中国居民人体营养健康状况趋势分析

下面定义一种评价指标函数对我国居民的人体营养健康趋势进行定量分析。首先应用连续区间有序加权平均 (COWA) 算子<sup>[28]</sup>将第  $j$  种营养素在第  $t$  年

的摄取标准区间  $NI_{jt} = [N_{jt}^-, N_{jt}^+]$  转化成实数为：

$$N_{jt} = \lambda N_{jt}^+ + (1 - \lambda) N_{jt}^- \quad (54)$$

对第  $j$  种营养素在第  $t$  年的人均每日摄入量  $A_{jt}$  和  $N_{jt}$  进行归一化处理：

$$A'_{jt} = \frac{N_{jt} - N_{jt}^-}{N_{jt}^+ - N_{jt}^-} \quad (55)$$

$$N'_{jt} = \frac{A_{jt} - N_{jt}^-}{N_{jt}^+ - N_{jt}^-} \quad (56)$$

则第  $t$  年的评价值定义为:

$$d_t = (\sum_j (A'_{jt} - N'_{jt})^2)^{\frac{1}{2}} \quad (57)$$

可以看出评价值越小说明营养总体的摄取越合理, 计算结果如图 84:

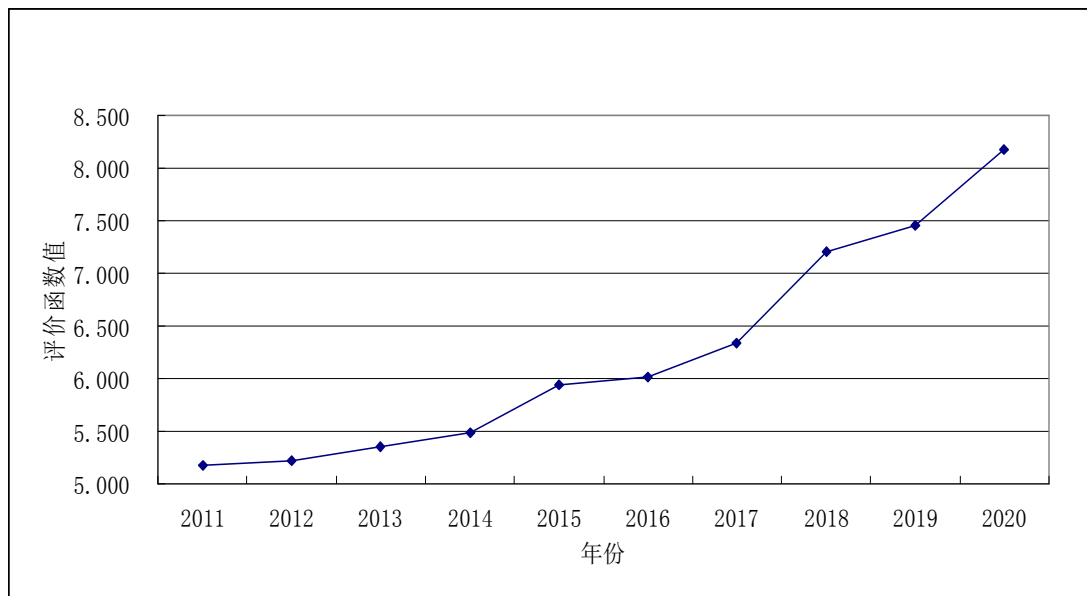


图 90 评价函数随时间变化趋势

从图 90 中可以看出随着时间的推移, 人体营养越来越不合理, 追究原因我们会发现, 造成这种现象的根本原因是维生素 A 的含量一直远远高于标准区间, 这是因为各种果蔬基本都含有它。而维生素 A 的过量摄入会导致各种疾病, 尤其是心肺功能和其它生理器官方面疾病。

而若将维生素 A 去除, 再计算出评价结果随时间的变化值如图 91。

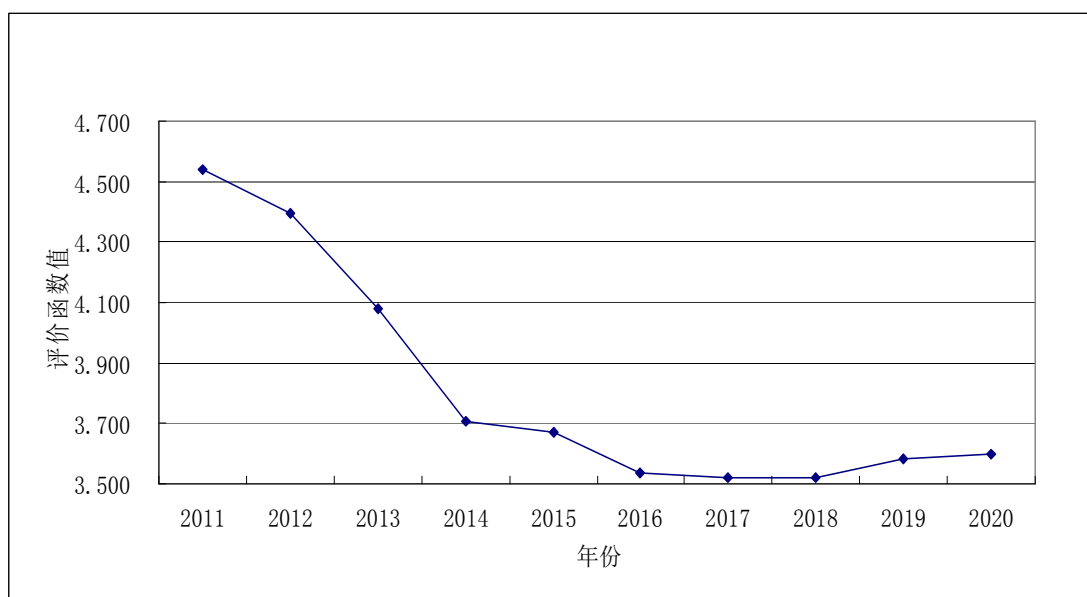


图 91 去除维生素 A 评价函数随时间变化趋势

从图 91 中我们可以看出: 评价函数值从 2011 年到 2018 年呈下降趋势, 2019 年和 2020 年有所回升。这说明人体营养的摄取直到 2018 年是越来越

越合理，但是在 2019 年和 2020 年会有呈现不合理的趋势。这是因为现阶段我国还是发展中国家，各种果蔬的摄入量都是呈上升趋势，用上升趋势预测的结果还是上升趋势，随着果蔬摄入量越来越多，总会出现像 2019 年和 2020 年这种不合理现象。因此，都可以看出我国居民的人体营养健康状况将会恶化。

4.3 问题三模型的建立和求解

4.3.1 水果与蔬菜的不可替代分析

根据问题二公式（52）可以计算出在每一年水果和蔬菜对人体各种营养素提供量，结果如图 92—101。

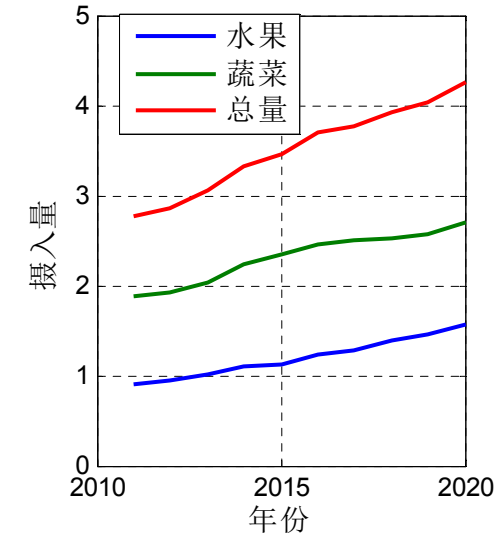


图 92 水果和蔬菜对锌提供量

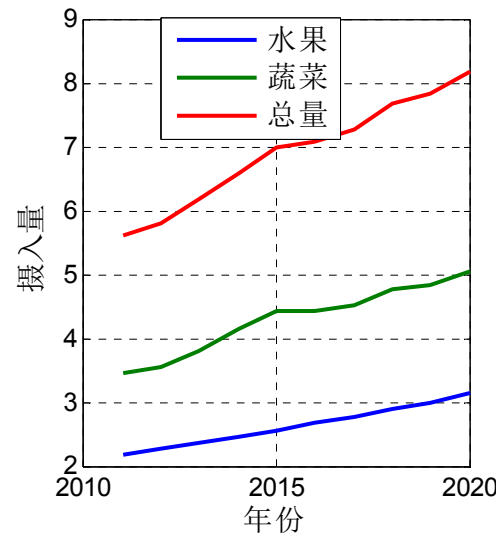


图 93 水果和蔬菜对维生素 E 提供量

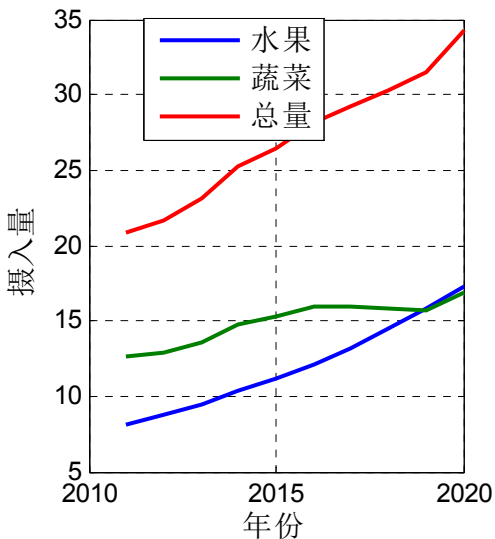


图 94 水果和蔬菜对维生素 C 提供量

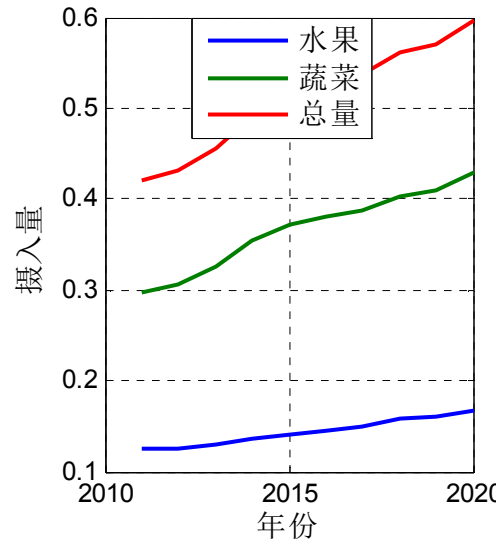


图 95 水果和蔬菜对维生素 B2 提供量

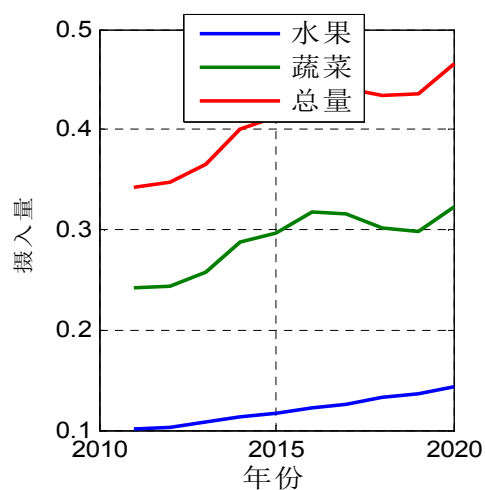


图 96 水果和蔬菜对维生素 B1 提供量

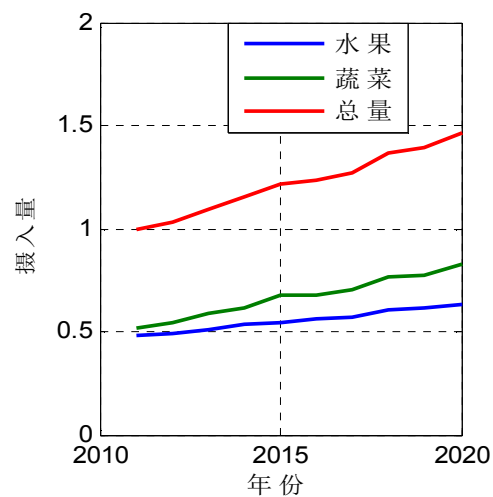


图 97 水果和蔬菜对维生素 A 提供量

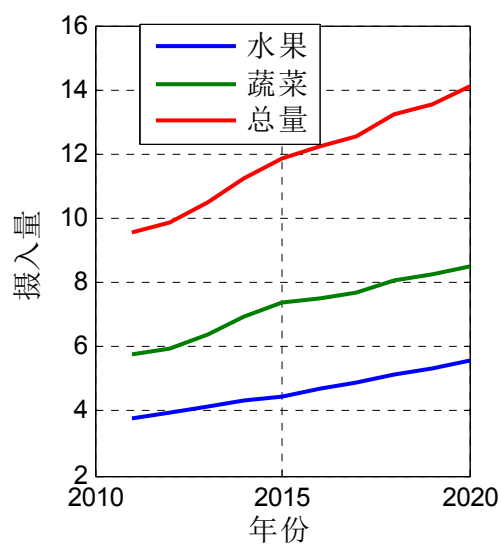


图 98 水果和蔬菜对膳食纤维提供量

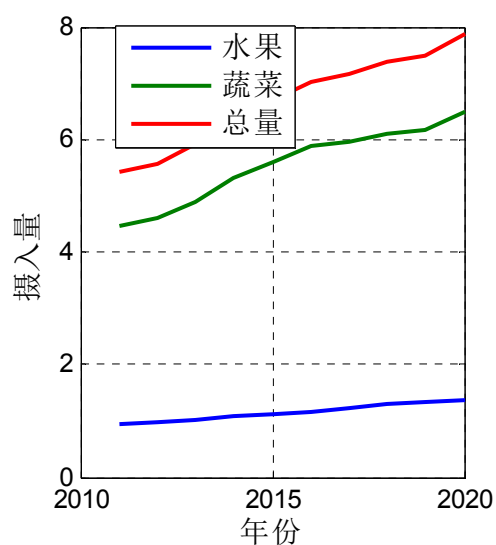


图 99 水果和蔬菜对铁提供量

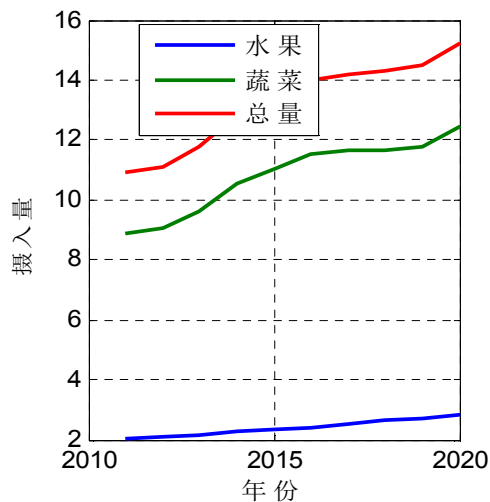


图 100 水果和蔬菜对蛋白质提供量

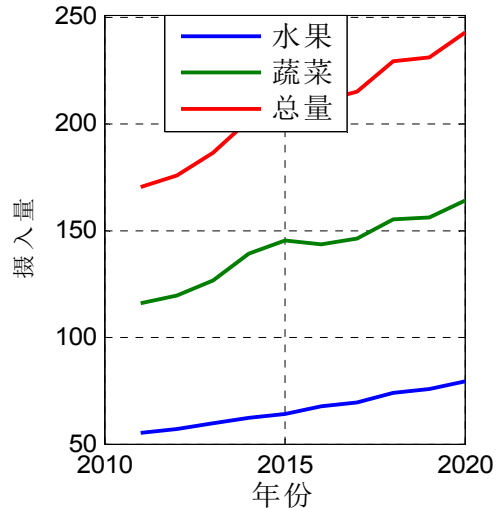


图 101 水果和蔬菜对钙提供量

从图 92—图 101 可以看出，蔬菜对锌、维生素 E、维生素 C、维生素 B2、维生素 B1、维生素 A、膳食纤维、铁、蛋白质、钙等营养素提供量基本大于水果的提供量。也就是说人体每年对这些营养素的摄取绝大部分是来源于蔬菜，而往往少量蔬菜就可以补充的营养量，要需要大量水果来补充，这样即不经济也不健康，所以用水果替代蔬菜不可取；而含有维生素 B6 等营养素的只有水果，所以我们也不能用蔬菜替代水果。综上所述，在这里我们只考虑蔬菜与蔬菜、水果与水果之间可以相互替代，而忽略蔬菜与水果之间的替代。

#### 4.3.2 应用自组织特征映射（SOM）的聚类分析<sup>[29-30]</sup>对水果、蔬菜进行可替代聚类

##### SOM网络模型

SOM神经网络模型由处理单元、网络拓扑结构和学习规则组成。处理单元是神经网络的基本单元，用以模拟生物神经元的功能。一个处理单元有多个输入、输出路径，输入端模拟脑神经元的树突功能，起信息传递作用。输出端模拟脑神经元的轴突功能，将信息处理结果向外界输出。

常用的SOM神经网络模型由输入层和输出层组成。输入层各神经元通过权向量将外界信息汇集到输出层的各神经元。输出层节点与输入层节点为全连接，输入层、输出层内部节点间没有连接关系。输出层内的每个神经元与其邻域连接，

此连接是相互激励的关系，训练后输出层不同节点代表不同的分类模式，所以SOM的输出层也称特征映射层。SOM神经网络拓扑结构如图102所示。

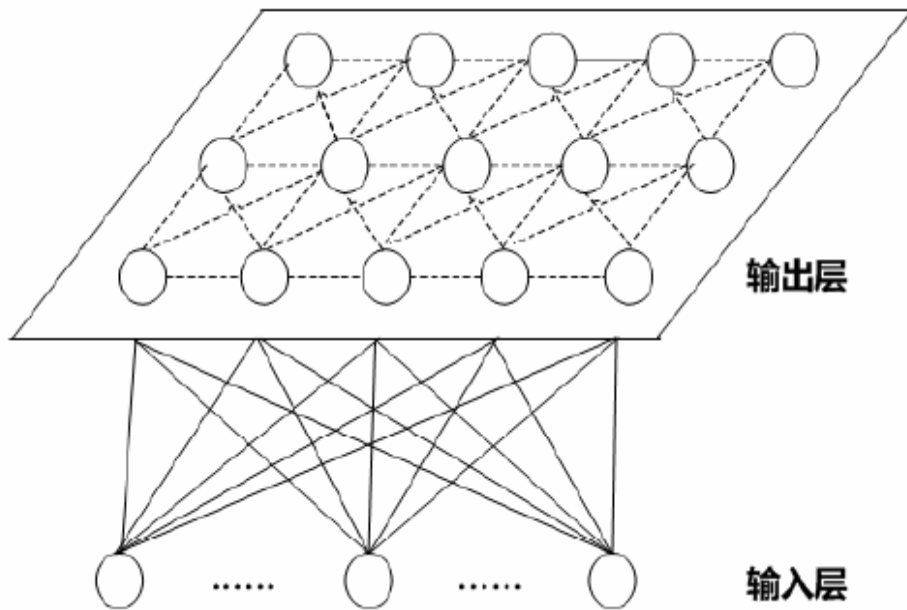


图102 SOM神经网络拓扑结构

自组织神经网络的目标，是在无监督的情况下，从输入数据中找出有意义的规律来。自组织人工神经网络以无教师教学的方式进行网络训练，网络通过自身训练，自动对输入模式进行分类。

#### SOM网络学习原理

SOM网络利用其自组织特性，将由N个输入模式组成的一维序列映射到二维神经元阵列上，通过自调整进行信息的特征提取或称聚类。这种自组织聚类过程是系统在自主、无导师指导的条件下完成的。SOM网络学习可分为以下两个主要过程：

##### 1) 神经元竞争学习过程

对于每一个输入向量，通过输入向量值与权重值之间的比较，在神经元之间产生竞争，权重向量与输入模式最相近的神经元被认为对于输入模式反映最强烈，将其标定为获胜的神经元，并称此神经元为输入模式的像，相同输入向量会在输出层产生相同的像。

设输入向量为  $X = [X_1, X_2, \dots, X_n]^T$ ，与输出层神经元  $j$  相应的权重向量  $w_j$  可表示为：

$$w_j = [w_{j1}, w_{j2}, \dots, w_{jp}]^T, \quad j = 1, 2, \dots, n \quad (58)$$

向量  $X$  的长度定义如下：

$$\|X\| = \left[ \sum_{i=1}^n x_i^2 \right]^{1/2} \quad (59)$$

在模式空间中两个向量  $X$  和  $Y$  之间的距离或长度表示如下：

$$\|X - Y\| = \left[ \sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2} \quad (60)$$

确定获胜输出层神经元，相当于选择权值向量  $w_j$  与输入向量  $X$  最为匹配的输层神经元，即选出权值向量与输入向量具有最小欧氏范数距离  $d_j$  的神经元作为获胜神经元。

$$d_j = \|X - w_j\| \quad (61)$$

## 2) 竞争层神经元的侧反馈过程

对于每个输入向量，会使竞争层相临近的神经元之间产生侧反馈，这种神经元之间的局部影响作用遵从以下规则：

- a) 以获胜神经元为圆心，对临近的神经元表现为兴奋性侧反馈。
- b) 以获胜神经元为圆心，对远邻的神经元表现为抑制性侧反馈。

这说明在竞争层，邻近神经元相互激励，远邻神经元相互抑制，比远邻更远的神经元则表现弱激励作用，通常，用“墨西哥帽”（图103）函数对神经元侧反馈进行计算。

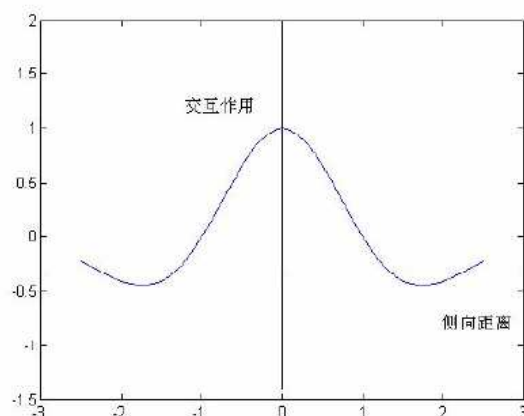


图103 墨西哥帽函数

侧反馈的作用是使得相近的特征能够在输出层阵列几何相近的位置上表现出来。应用侧反馈原理，在每一个获胜神经元附近形成一个“聚类区”。学习的结果总是使聚类区内各神经元的权重向量保持向输入向量逼近的趋势，从而使具有相近特性的输入向量聚集在一起，这个过程就是自组织。侧反馈的调整可以通过简单地调整邻域的大小来体现：正向反馈越大，负向反馈就越小。

邻域函数  $h_{j,i(x)}(n)$  定义了围绕获胜神经元邻近区域的大小。  $h_{j,i(x)}(n)$  是一个迭代函数，使用这一函数侧反馈的大小可在整个训练网络的过程中变化。邻域越大意味着正向反馈越多，训练区域越大。随后邻域的缩小使类别更小，因此类别也分得更细。半径为零时，仅仅包含获胜神经元。

### SOM算法描述：

Kohonen的SOM 算法<sup>[38]</sup>的本质是它用一个简单的几何计算代替类Hebb规则的复杂性质和侧向相互作用。算法的主要构成有：根据一定概率分布产生激活模式的连续输入空间，建立初始优胜邻域，设置学习率初值；以神经元的网格形式表示的网络拓扑，它定义一个离散输出空间；在获胜神经元周围定义随时间变化的邻域函数；学习率参数随时间递减，但永不为零。

算法具体步骤如下：

### 步骤1：初始化

对初始权值向量  $w_j(0)$  选择随机值。这里惟一的限制是对  $j = 1, 2, \dots, l$ ， $w_j(0)$  互不相同，其中  $l$  是网络中神经元的数目。可能希望保持较小的权值。还有一种算法初始化方法，是从输入向量  $\{x\}_{i=1}^N$  的可用集里随机选择权值向量  $\{w_j(0)\}_{j=1}^l$ 。建立初始优胜邻域  $h_{j_i}(0)$ ，对学习率赋初始值  $\eta(0)$ 。

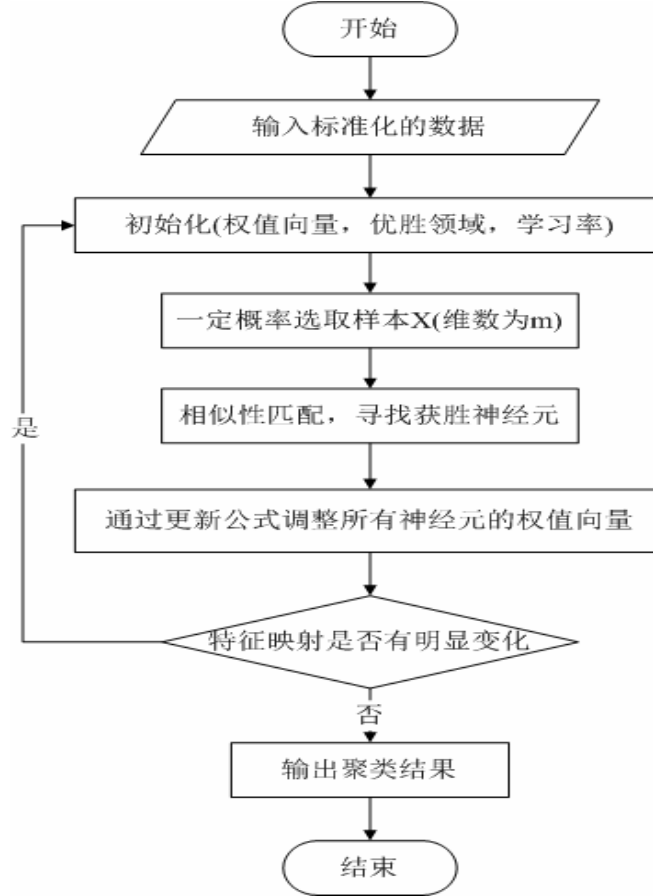


图104 Kohonen 算法流程图

### 步骤2：取样

以一定概率从输入空间取样本  $X$ ；向量  $X$  表示应用于网络的激活模式，向量  $X$  的维数等于  $m$ 。

### 步骤3：相似性匹配

在时间步  $n$  使用最小欧式距离准则寻找最匹配（获胜）的神经元  $i(x)$ ：

$$i(x) = \arg \min_j \|x(n) - w_j\|, \quad j = 1, 2, \dots, l \quad (62)$$

### 步骤4：更新

通过用更新公式

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \quad (63)$$

调整所有神经元的权值向量，其中  $\eta(n)$  是学习率参数， $h_{j,i(x)}(n)$  是获胜神经



元  $i(x)$  周围的邻域函数；为了获得更好的结果， $\eta(n)$  和  $\eta(n)$  在学习过程中是动态变化的。

**步骤5：**继续

继续步骤2直到在特征映射里观察不到明显的变化为止。

网络输入层神经元个数分别为蔬菜和水果的个数，训练前，为消除不同量纲和数值差异造成的误差，需对原始数据进行标准化模式预处理。输出层设置二维平面空间，由 6 个神经元节点构成（2×3）。初始学习速率  $\eta_0$  取 0.1，最大训练次数为 1000 次。

对水果和蔬菜求解结果为：

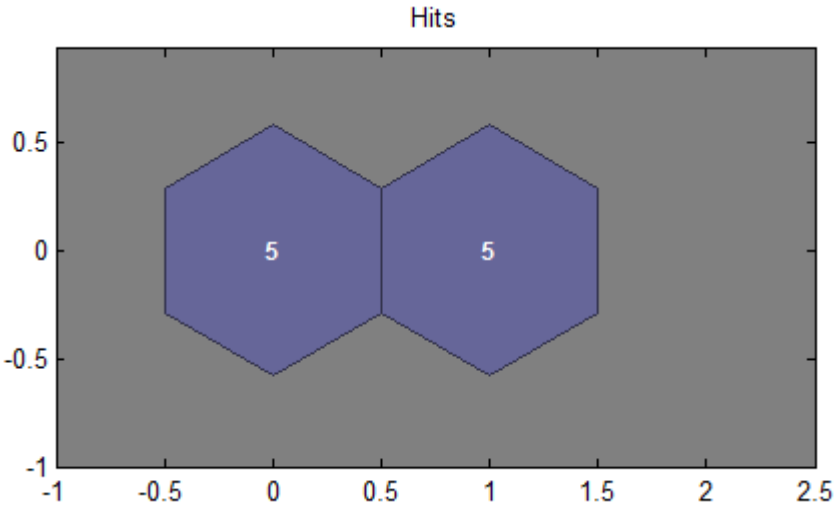


图 105 水果每类中品种个数

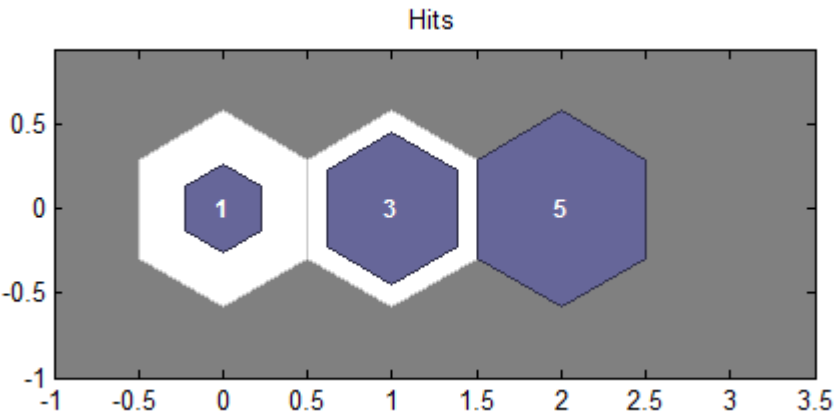


图 106 蔬菜每类中品种个数

对水果的聚类结果为：类别一：苹果、梨、桃、猕猴桃、李子；类别二：柑橘、香蕉、葡萄、红枣、西瓜。

对蔬菜的聚类结果为：类别一：蘑菇；类别二：土豆、大白菜、菠菜；类别三：西红柿、茄子、胡萝卜、青椒、菜花；

每种类别可以看成所含营养素大致一样，可以相互替代。类别之间的营养素不能相互替代，也就是说每个季节我们必须消费每种类别中的一个或多个

品种，这样还可以保证饮食多样性。

### 4.3.3 蔬菜和水果人均消费量最优化模型建立

因为水果与蔬菜不可相互替代，所以我们分别对水果和蔬菜建立最优化模型。这里需要对各种营养素标准按照比例分配给水果和蔬菜。

设  $NI_j = [N_j^-, N_j^+]$  为第  $j$  种营养素的摄取量标准区间，设  $f_j$  和  $v_j$  分别为水果、蔬菜对该营养素的贡献量， $f_j$ 、 $v_j$  的计算参考图（92-101）。则：第  $j$  种营养素的摄取量标准区间（水果提供）为：

$$NIF_j = [NF_j^-, NF_j^+] \quad (64)$$

其中  $NF_j^- = f_j N_j^-$ ,  $NF_j^+ = f_j N_j^+$ ；

第  $j$  种营养素的摄取量标准区间（蔬菜提供）为：

$$NIV_j = [NV_j^-, NV_j^+] \quad (65)$$

其中  $NV_j^- = v_j N_j^-$ ,  $NV_j^+ = v_j N_j^+$ 。

在我国由于南北气候不同，会导致果蔬在南北种植分布不同，从而影响其价格南北分差较大。各可替代类别果蔬的价格如图所示。

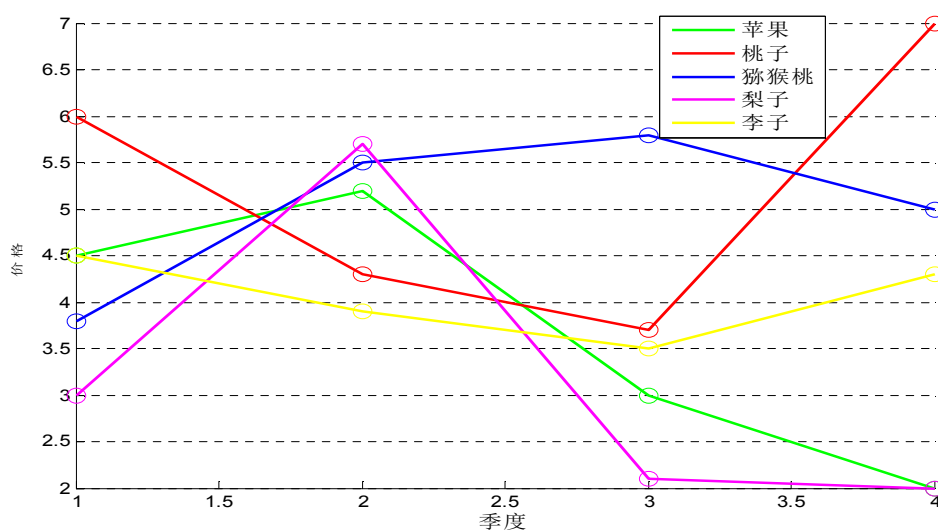


图 107 可互相替代类别一水果的北方价格

由图 107 可以看出北部区域：在第一季度梨子的价格最低可以用来替换其它水果；第二季度李子的价格变为最低可以替换其他水果；第三季度梨子的价格最低可以替换其他水果；第四季度梨子和苹果的价格最低可以用来替换其他水果。

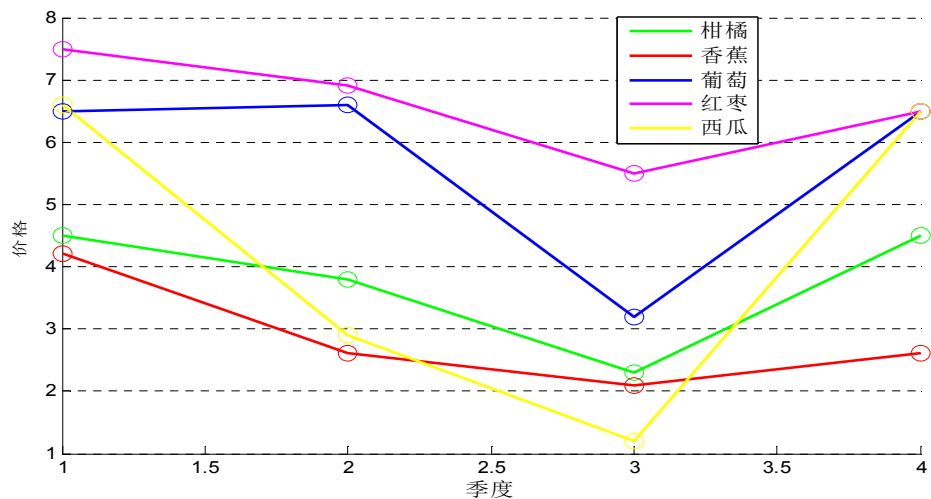


图 108 可互相替代类别二水果的北方价格

由图 108 可以看出北部区域：在第一季度香蕉的价格最低可以用来替换其它水果；第二季度还是香蕉的价格最低可以替换其他水果；第三季度西瓜的价格最低可以替换其他水果；第四季度回到香蕉的价格最低可以用来替换其他水果。

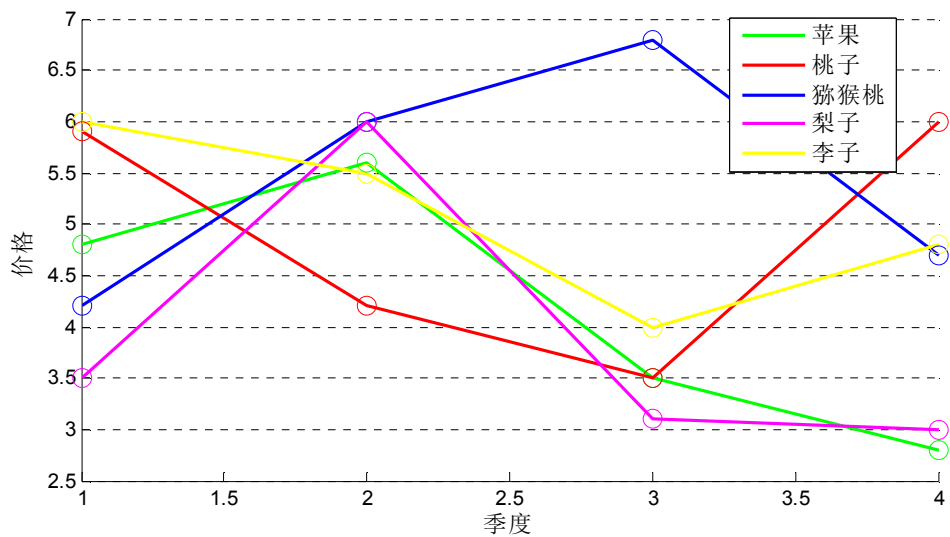


图 109 可互相替代类别一水果的南方价格

由图 109 可以看出南部区域：在第一季度梨子的价格最低可以用来替换其它水果；第二季度桃子的价格变为最低可以替换其他水果；第三季度梨子的价格最低可以替换其他水果；第四季度苹果的价格最低可以用来替换其他水果。

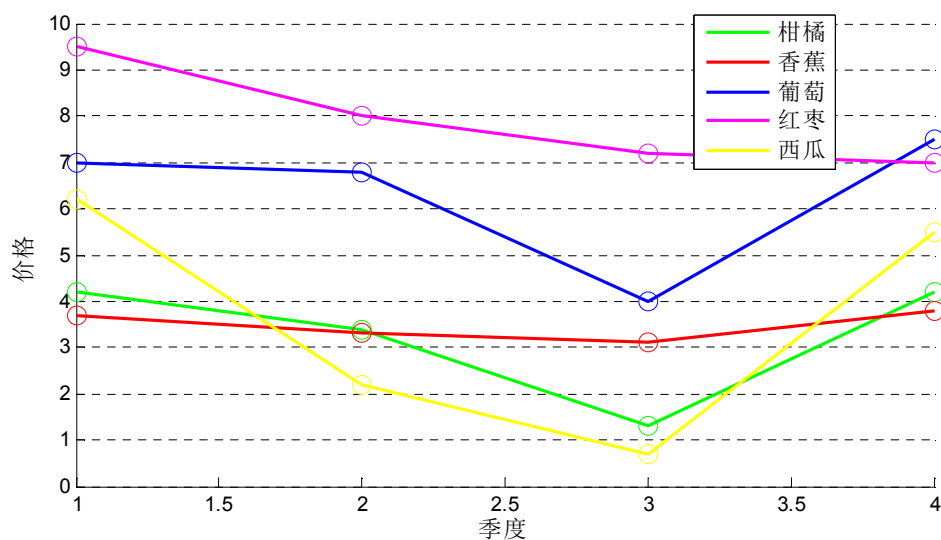


图 110 可互相替代类别二水果的南方价格

图 110 可以看出南部地区：在第一季度香蕉的价格最低可以用来替代其它水果；第二季度西瓜的价格变为最低可以替换其他水果；第三季度西瓜的价格最低可以替换其他水果；第四季度香蕉的价格最低可以用来替换其他水果。

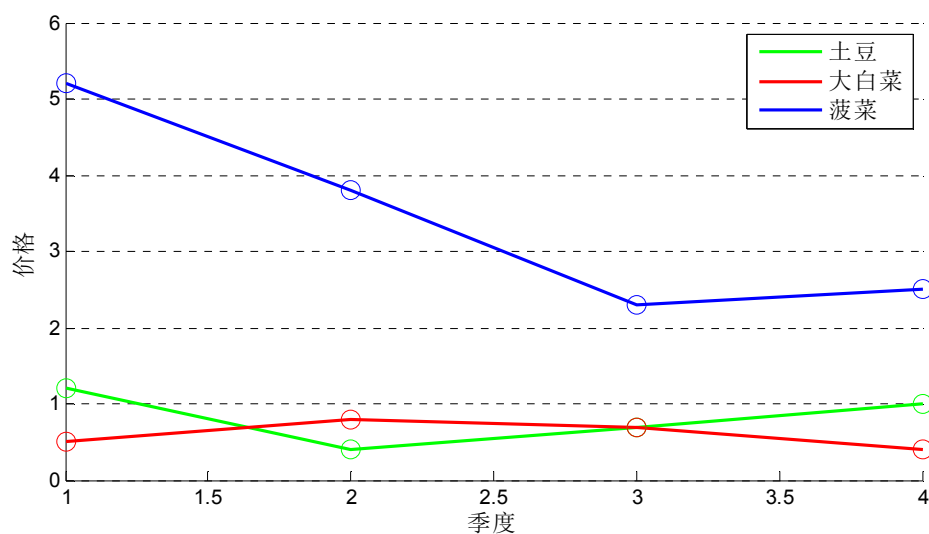


图 111 可互相替代类别一蔬菜的北方价格

图 111 可以看出北部地区：在第一季度大白菜的价格最低可以用来替代其它蔬菜；第二季度土豆的价格变为最低可以替换其他蔬菜；第三季度土豆和大白菜的价格最低可以替换其他蔬菜；第四季度大白菜的价格最低可以用来替换其他蔬菜。

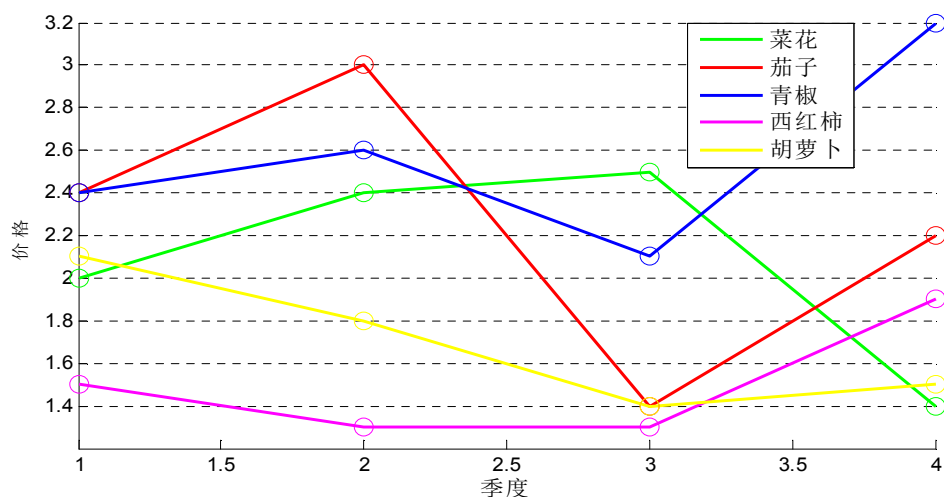


图 112 可互相替代类别二蔬菜的北方价格

图 112 可以看出北部地区：在第一季度西红柿的价格最低可以用来替代其它蔬菜；第二季度西红柿的价格变为最低可以替换其他蔬菜；第三季度西红柿的价格最低可以替换其他蔬菜；第四季度菜花的价格最低可以用来替换其他蔬菜。

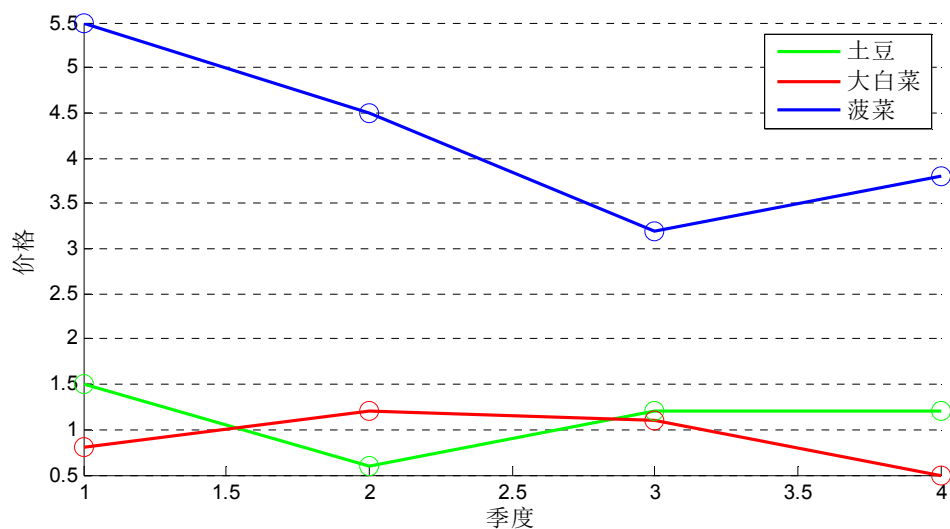


图 113 可互相替代类别一蔬菜的南方价格

图 113 可以看出南部地区：在第一季度大白菜的价格最低可以用来替代其它蔬菜；第二季度土豆的价格变为最低可以替换其他蔬菜；第三季度大白菜的价格最低可以替换其他蔬菜；第四季度大白菜的价格最低可以用来替换其他蔬菜。

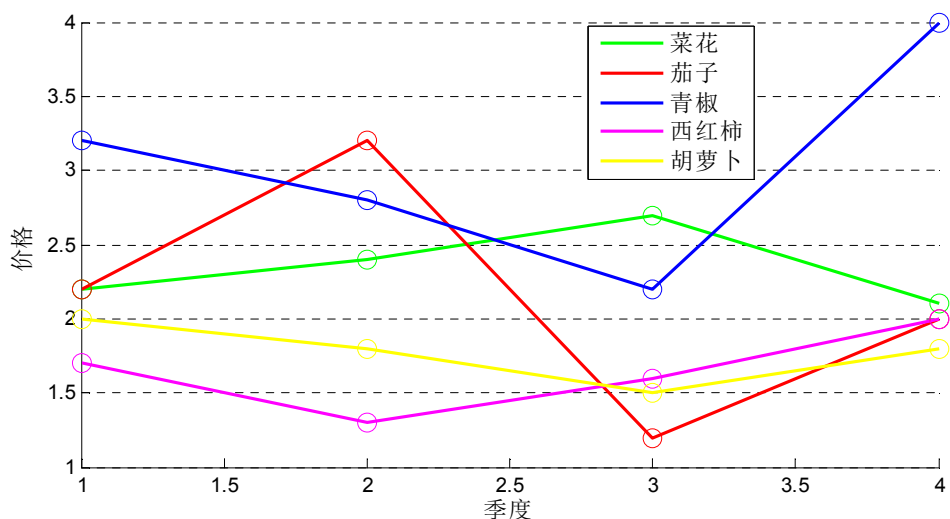


图 114 可互相替代类别二蔬菜的南方价格

图 114 可以看出南部地区：在第一季度西红柿的价格最低可以用来替代其它蔬菜；第二季度西红柿的价格变为最低可以替换其他蔬菜；第三季度茄子的价格最低可以替换其他蔬菜；第四季度胡萝卜的价格最低可以用来替换其他蔬菜。

#### (1) 对水果建立最优化模型

设苹果、梨、桃、猕猴桃、李子、柑橘、香蕉、葡萄、红枣、西瓜分别为第  $i(i=1,2,\dots,10)$  种水果。则前五种可以相互替代，第 6 到 10 种水果可以相互替代。

设  $PF_{ik}$ 、 $PF'_{ik}$  分别为第  $i(i=1,2,\dots,10)$  种水果在第  $k(k=1,2,3,4)$  季度的南方和北方的价格， $x_{ik}$ 、 $x'_{ik}$  分别为第  $i$  种水果在第  $k$  季度的南方和北方消费量，可以算出水果年消费价格为。作为消费者，我们希望消费价格越低越好，因此目标函数为：

$$\min S = \sum_i \sum_k (PF_{ik} x_{ik} + PF'_{ik} x'_{ik}) \quad (66)$$

设  $C_i$  为第  $i$  种水果的年总消费量，因为四个季度南北消费量总和不能超过总消费量，所以可以得到约束条件一：

$$\begin{cases} \sum_k (x_{ik} + x'_{ik}) \leq C_i \\ x_{ik} + x'_{ik} > 0 \end{cases} \quad (67)$$

设  $a_{ij}$  为第  $i$  种水果的第  $j$  种营养素的含量， $P$  为总人口数，我们需要保证人均每种营养素的摄入量要达到标准营养区间，所以可以得到约束条件二：

$$\begin{cases} \frac{\sum_k (\sum_i (x_{ik} + x'_{ik}) a_{ij})}{365P} \leq NF_j^+ \\ \frac{\sum_k (\sum_i (x_{ik} + x'_{ik}) a_{ij})}{365P} \geq NF_j^- \end{cases} \quad (68)$$

因为把水果分为了两个类别，每个类别内的营养素含量大致相同，所以每个类别内的水果在每个季节必须吃一个品种，以保证营养多样性和饮食多样性，这样也实现了水果每个类别内部在各个季度替代。所以可以得到约束条件三：

$$\begin{cases} \sum_{i=1}^5 x_{ik} > 0 \\ \sum_{i=1}^5 x'_{ik} > 0 \\ \sum_{i=6}^{10} x_{ik} > 0 \\ \sum_{i=6}^{10} x'_{ik} > 0 \end{cases} \quad (69)$$

运用问题一的 PSO-GA 算法求解得：

表 10 北方各个季度水果的消费量

		第一季度	第二季度	第三季度	第四季度
猕猴桃	类别一	16802	7203	1360	3757
李子		12	11	9603	8638
梨		8399	0	6201	8994
桃		0	0	3360	0
苹果		19197	22396	29402	25198
柑橘类	类别二	0	6812	13778	0
香蕉		2964	3961	3723	3802
葡萄		0	59402	33281	2596
红枣		15	13801	16504	4810
西瓜		0	52204	33600	26010

从表中可以看出南方各个季度对各种水果的摄入量：在第一季度我们主要吃类别一中的猕猴桃、桃子，类别二中的香蕉；在第二季度中类别二中的水果都有摄入，而类别一种只选择吃猕猴桃和香蕉；在第三季度是各种水果收获的季节，所以所有水果都有所摄入；在第四季度除了类别一中的桃子，类别二中的柑橘类，其他水果均可以吃。

表 11 北方各个季度水果的消费量

		第一季度	第二季度	第三季度	第四季度
猕猴桃	类别一	12920	5499	1160	2300
李子		18	15.6	8420	7740
梨		7213	22.8	6300	6806
桃		0	0	3552	0
苹果		22512	24960	2684	20800

柑橘类	类别二	0	7600	17480	0
香蕉		3360	2340	1890	2080
葡萄		0	69960	38402	0
红枣		30	13789	16500	5200
西瓜		0	46400	28800	7814

从表中可以看出北方各个季度对各种水果的摄入量：在第一季度我们主要吃类别一中的猕猴桃、桃子，类别二中的香蕉；在第二季度中类别二中的水果都有摄入，而类别一种只选择吃猕猴桃和香蕉；在第三季度是各种水果收获的季节，所以所有水果都有所摄入；在第四季度除了类别一中的桃子，类别二中的柑橘类和葡萄，其他水果均可以吃。

## (2) 对蔬菜建立最优化模型

设蘑菇、土豆、大白菜、菠菜、西红柿、茄子、胡萝卜、青椒、菜花分别为第  $i(i=1,2\cdots 9)$  种蔬菜； $PV_{ik}$ 、 $PV'_{ik}$  分别为第  $i(i=1,2\cdots 9)$  种蔬菜在第  $k(k=1,2,3,4)$  季度的南方和北方的价格； $x_{ik}$ 、 $x'_{ik}$  分别为第  $i$  种蔬菜在第  $k$  季度的南方和北方消费量；设  $C_i$  为第  $i$  种蔬菜的年总消费量； $a_{ij}$  为第  $i$  种蔬菜的第  $j$  种营养素的含量。

同理对水果建立最优化模型可以得到对蔬菜建立最优化模型为：

$$\begin{aligned}
 \min \quad & S = \sum_i \sum_k (PV_{ik} x_{ik} + PV'_{ik} x'_{ik}) \\
 s.t. \quad & \begin{cases} \sum_k (x_{ik} + x'_{ik}) \leq C_i \\ x_{ik} + x'_{ik} > 0 \\ \frac{\sum_k (\sum_i (x_{ik} + x'_{ik}) a_{ij})}{365P} \leq NV_j^+ \\ \frac{\sum_k (\sum_i (x_{ik} + x'_{ik}) a_{ij})}{365P} \geq NV_j^- \\ x_{1k} > 0 \\ x'_{1k} > 0 \\ \sum_{i=2}^4 x_{ik} > 0 \\ \sum_{i=2}^4 x'_{ik} > 0 \\ \sum_{i=5}^9 x_{ik} > 0 \\ \sum_{i=5}^9 x'_{ik} > 0 \end{cases}
 \end{aligned} \tag{70}$$

运用问题一的 PSO-GA 算法求解得：

表 12 南方各个季度蔬菜的消费量



		第一季度	第二季度	第三季度	第四季度
土豆	类别二	21020	11994	16821	16800
大白菜		3201	240	0	3510
菠菜		11012	18010	25612	22814
西红柿	类别三	4080	11180	12160	4800
茄子		17600	12793	15600	16012
胡萝卜		3198	6480	6598	7200
青椒		7680	7840	8800	7984
菜花		3523	2880	1080	4200

表 13 北方各个季度蔬菜的消费量

		第一季度	第二季度	第三季度	第四季度
土豆	类别二	16800	8012	9800	13987
大白菜		2478	320	0	3200
菠菜		11440	15213	17020	14025
西红柿	类别三	2974	10400	8580	3800
茄子		16800	1182	16800	17601
胡萝卜		3360	6480	6160	3120
青椒		5760	7279	8400	6400
菜花		3224	3120	1496	3360

从表 12--13 中可以看出南方和北方各个季度对各种蔬菜的摄入趋势大体一致：在第一季度我们主要吃类别二中的土豆、大白菜，类别三中的茄子，其它蔬菜均有摄入；在第二季度我们主要吃类别二中的土豆、菠菜，类别三中的西红柿和茄子，其它蔬菜均有摄入；在第三季度我们主要吃类别二中的土豆、菠菜，类别三中的西红柿和茄子，其它蔬菜除了大白菜均有摄入；在第四季度对各种蔬菜均基本都是大量摄入。

#### 4.4 问题四模型的建立与求解

##### 4.4.1 基于果蔬人均消费量的多目标规划模型的建立

设决策变量  $x_i$  为第  $i$  种果蔬年度合理人均消费量， $S_1$  为果蔬在田间地头到大市场损耗率， $S_2$  为大市场到零售市场损耗率， $IM_i$  为第  $i$  中果蔬的进口量， $EX_i$  为第  $i$  中果蔬的出口量， $P$  为人口数，则第  $i$  种果蔬年度用来销售的生产量为：

$$O'_i = \frac{Px_i + EX_i - IM_i}{(1 - S_1)(1 - S_2)}$$

设  $O_i$  为年度实际生产值，他一定不能小于用来销售的生产量，所以我们得到约束条件一：

$$\frac{Px_i + EX_i - IM_i}{(1 - S_1)(1 - S_2)} \leq O_i$$

如果用来销售的生产量小于实际生产值，又会使部分果蔬滞销，造成种植

者利益受损。若要使种植者能够尽量获得较大收益，则需要使滞销的果蔬的价格总和最小，于是有目标函数一：

$$\min \sum_i (O_i - O'_i) P_i$$

其中  $P_i$  为第  $i$  种果蔬的价格，同时我们也要顾虑居民的购买成本，也要顾及居民的购买成本，使其购买成本达到最小，于是又得到目标函数二：

$$\min P_i x_i$$

设  $a_{ij}$  表示第  $i$  种果蔬品种第  $j$  种营养素的含量，则需要控制营养摄入量使其达到标准范围，应用问题一公式（）有约束条件二：

$$u_j \left( \sum_i \frac{x_i a_{ij}}{365} \right) > \theta$$

因为目标函数均为达到极小，所以我们可以得到如下最优化模型：

$$\begin{aligned} \min & \sum_i (O_i - O'_i) P_i + \varphi P_i x_i \\ \text{s.t.} & \begin{cases} \frac{P x_i + EX_i - IM_i}{(1 - S_1)(1 - S_2)} \leq O_i \\ u_j \left( \sum_i \frac{x_i a_{ij}}{365} \right) > \theta \end{cases} \end{aligned}$$

其中  $\varphi$  为加权系数。

#### 4.4.2 基于果蔬人均消费量规划模型的求解

用问题一的 PSO-GA 算法求解得结果如图 115—116：

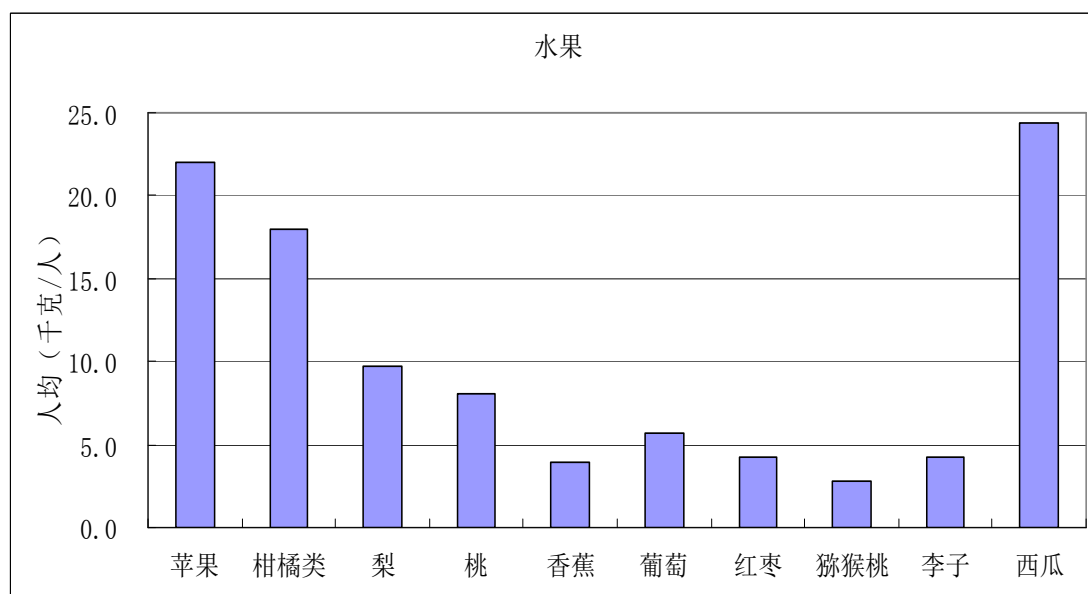
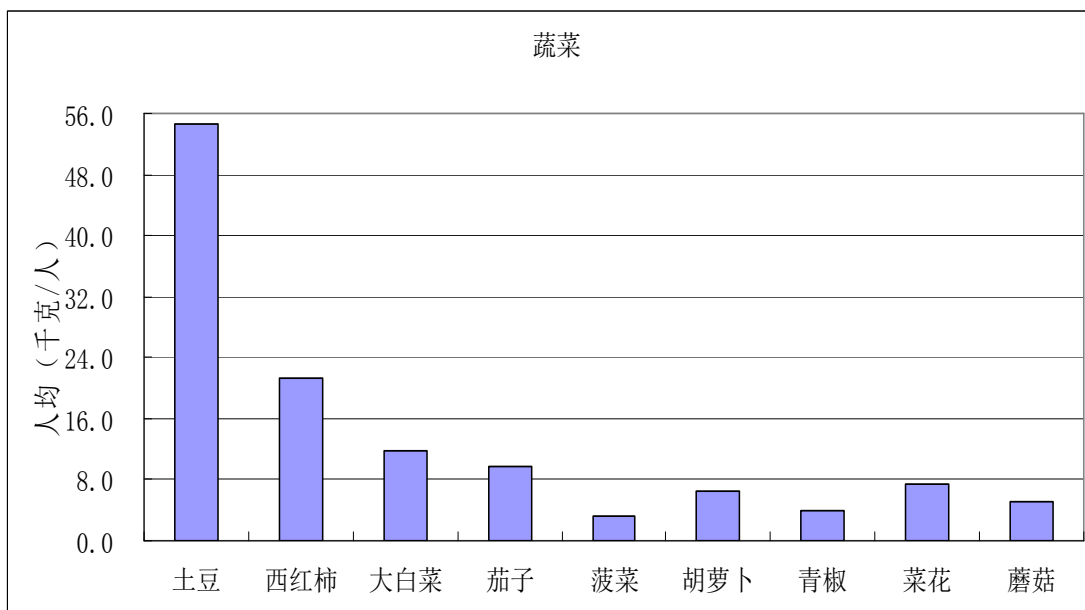


图 115 人均水果年消费量



116 人均蔬菜年消费量

同时，我们得到与问题一的人均消费量实际值对比如图：

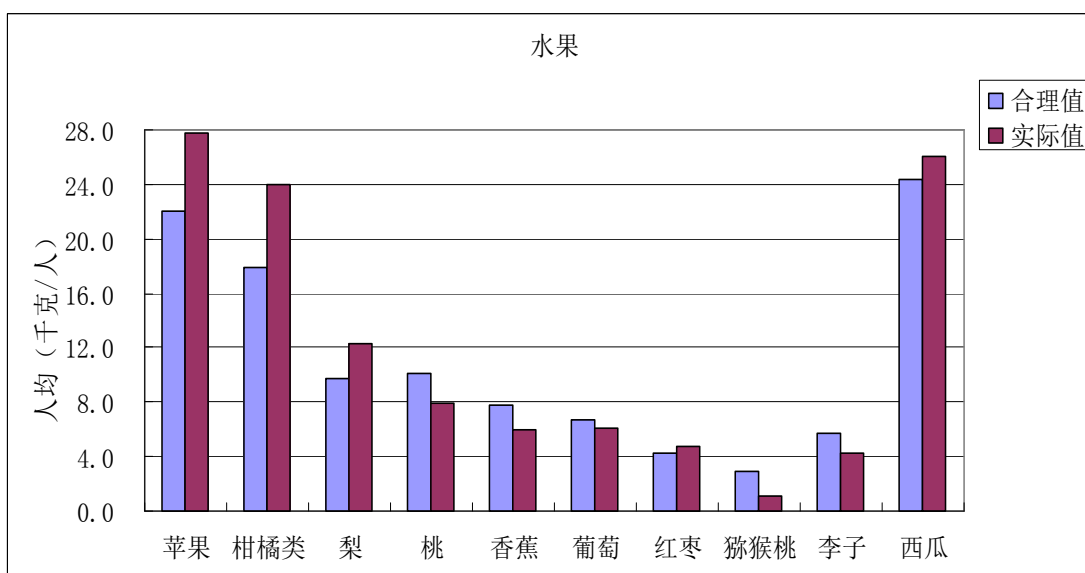


图 117 水果人均消费量的实际值与合理值

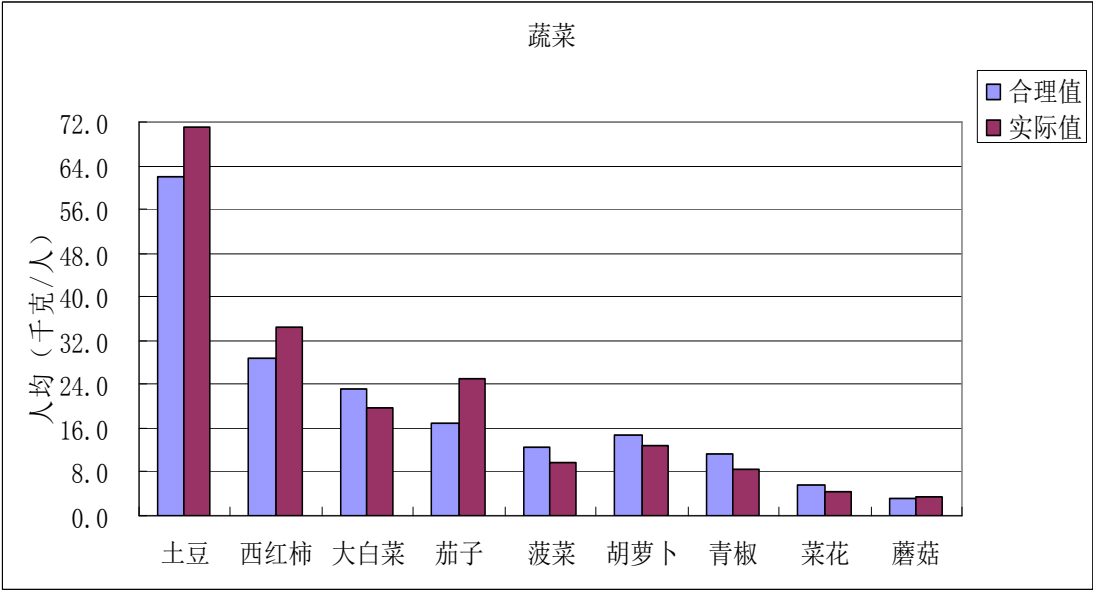


图 118 蔬菜人均消费量的实际值与合理值

可以看出除了苹果、柑橘类、梨、西瓜、土豆、西红柿、茄子的人均消费值高于合理值外，其它基本小于合理值。对照营养素表可以发现，因为这些果蔬的维生素 A 和维生素 E 的含量较高，而根据第二问中我国健康状况分析中人均摄入维生素 A 和维生素 E 高于标准值相符合。因此在今后的产量规模的战略调整是：适当减少苹果、柑橘类、梨、西瓜、土豆、西红柿和茄子的种植面积，加大对桃、香蕉、葡萄、猕猴桃、李子、菠菜、胡萝卜、青椒、菜花的种植。

另外，我们得到主要水果和蔬菜一直到 2020 及以后几年的产量、进口量、出口量和种植面积，结果如表 14—21 所示。

表 14 水果产量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
苹果	4236.50	4368.62	4500.74	4632.86	4764.98	4873.98	5020.37	5152.49	5284.61	5416.73
柑橘类	2937.69	3069.81	3201.93	3334.05	3466.17	3575.17	3707.29	3839.41	3971.53	4103.65
梨	1849.61	1829.61	1961.73	1969.17	2101.29	2210.29	2342.41	2474.53	2606.65	2738.77
桃	1121.19	1253.31	1385.43	1517.55	1649.67	1758.67	1890.79	2022.91	2155.03	2287.15
香蕉	1032.24	1164.36	1296.48	1428.60	1560.72	1669.72	1801.84	1933.96	2066.08	2198.20
葡萄	968.97	1053.09	1137.21	1221.33	1305.45	1389.57	1473.69	1557.81	1641.93	1726.05
红枣	654.86	696.98	739.10	781.22	823.34	865.46	907.58	949.70	991.82	1033.94
猕猴桃	147.17	179.29	211.41	243.53	275.65	307.77	339.89	372.01	404.13	436.25
李子	612.09	694.21	776.33	858.45	940.57	1022.69	1104.81	1186.93	1269.05	1351.17
西瓜	7214.83	7346.95	7479.07	7611.19	7743.31	7852.31	7984.43	8116.55	8248.67	8380.79

表 15 蔬菜产量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
土豆	9259.61	9448.09	9636.57	9825.05	10013.53	10202.01	10390.49	10578.97	10767.45	10955.93
西红柿	5813.62	5950.10	6086.58	6223.06	6359.54	6496.02	6632.50	6768.98	6905.46	7041.94
大白菜	3351.25	3459.73	3568.21	3676.69	3785.17	3893.65	4002.13	4110.61	4219.09	4327.57
茄子	3218.01	3326.49	3434.97	3543.45	3651.93	3760.41	3868.89	3977.37	4085.85	4194.33
菠菜	2043.12	2151.60	2260.08	2368.56	2477.04	2585.52	2694.00	2802.48	2910.96	3019.44

胡萝卜	1846.20	1954.68	2063.16	2171.64	2280.12	2388.60	2497.08	2605.56	2714.04	2822.52
青椒	1734.36	1842.84	1951.32	2059.80	2168.28	2276.76	2385.24	2493.72	2602.20	2710.68
菜花	938.91	1007.39	1075.87	1144.35	1212.83	1281.31	1349.79	1418.27	1486.75	1555.23
蘑菇	587.98	696.46	804.94	913.42	1021.90	1130.38	1238.86	1347.34	1455.82	1564.30

表 16 水果进口量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
苹果	30.39	31.87	34.80	28.85	30.85	54.70	36.05	84.25	37.06	69.83
香蕉	47.26	41.06	48.77	44.93	42.92	46.85	41.24	44.78	57.87	73.85
红枣	1.02	1.02	1.09	1.14	1.32	1.31	0.80	0.88	1.11	1.54
葡萄	20.22	25.98	24.03	25.76	27.41	25.44	24.24	26.71	33.91	33.69
柑橘类	49.55	55.16	63.32	55.17	58.93	63.26	57.32	53.49	59.89	68.00
梨	4.45	4.27	4.40	4.96	4.70	4.26	4.94	4.20	3.86	3.65
桃	5.73	5.87	4.75	4.80	4.55	4.29	4.14	3.95	3.31	3.68
猕猴桃	2.49	2.51	2.38	2.81	3.60	4.40	5.10	5.64	6.85	7.57
李子	5.24	4.94	5.80	3.30	4.60	3.74	3.86	3.80	4.18	5.05
西瓜	8.86	12.34	9.06	11.52	19.91	21.42	22.86	26.22	30.07	34.54

表 17 水果出口量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
苹果	67.34	95.17	201.87	151.31	178.67	218.69	339.52	354.39	247.50	297.24
香蕉	4.29	4.56	5.96	5.35	5.37	5.81	5.50	4.03	4.50	3.70
红枣	0.14	0.08	0.09	0.04	0.05	0.07	0.06	0.06	0.05	0.06
葡萄	6.45	9.84	11.14	13.23	14.77	18.45	21.15	24.69	36.61	33.69
柑橘类	24.25	29.60	38.06	43.55	50.68	49.06	56.85	90.42	117.79	102.34
梨	18.55	2.45	29.92	31.97	36.99	38.01	40.90	45.00	46.40	43.86
桃	3.77	1.24	1.92	1.57	1.76	2.09	2.50	2.68	4.09	2.83
猕猴桃	0.12	0.27	0.36	0.85	0.84	0.92	1.03	0.46	0.70	0.79
李子	1.03	1.83	1.76	1.24	1.57	0.85	1.79	1.80	2.92	2.84
西瓜	1.44	1.92	2.83	3.10	3.58	3.66	3.68	3.43	5.86	5.14

表 18 蔬菜进口量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
土豆	1.88	1.05	1.45	0.20	2.93	2.38	3.19	2.33	2.22	2.22
西红柿	0.27	0.32	0.26	0.73	0.77	0.80	0.94	0.97	0.91	0.87
大白菜	17.76	15.95	14.39	14.37	18.76	26.07	33.62	34.45	39.16	38.04
茄子	0.39	0.83	0.26	0.59	0.43	0.43	0.38	0.25	0.27	0.32
菠菜	0.52	0.60	0.39	0.22	0.13	0.76	0.59	0.67	0.20	0.27
胡萝卜	1.78	0.20	1.68	1.84	1.54	1.55	4.61	4.07	3.48	1.45
青椒	0.35	0.19	0.22	0.36	0.72	0.65	1.07	0.99	0.69	0.62
菜花	3.03	2.86	3.32	2.19	2.72	1.97	1.00	0.85	1.58	2.13
蘑菇	0.86	0.79	0.88	0.88	0.75	0.98	1.29	1.52	1.90	2.08

表 19 蔬菜出口量（单位：万吨）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
土豆	8.54	12.47	16.25	18.22	24.49	35.39	36.47	34.17	38.30	25.92
西红柿	2.83	3.38	5.88	8.29	8.60	7.67	8.98	11.67	10.81	10.41
大白菜	6.66	5.26	8.76	15.08	14.17	16.58	27.81	30.72	31.10	37.92
茄子	1.89	1.83	2.82	3.75	2.47	1.59	2.09	2.00	2.02	1.87
菠菜	0.83	0.61	1.22	1.42	1.95	1.81	1.04	0.75	0.62	0.75
胡萝卜	8.81	13.86	20.37	28.67	39.10	43.82	42.63	43.00	43.78	49.92
青椒	1.27	2.35	5.68	6.66	4.10	2.74	4.95	7.70	6.02	7.20
菜花	5.29	3.85	5.63	8.53	13.18	12.70	5.71	4.13	5.98	7.58
蘑菇	5.70	4.13	4.54	5.12	4.61	4.16	4.02	2.21	2.56	3.40

表 20 水果种植面积（单位：千公顷）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
苹果	2066.92	2079.08	2091.24	2103.40	2115.56	2127.72	2139.88	2152.04	2164.20	2176.36
香蕉	1428.27	1440.43	1452.59	1464.75	1476.91	1489.07	1501.23	1513.39	1525.55	1537.71
红枣	1086.05	1098.21	1110.37	1122.53	1134.69	1146.85	1159.01	1171.17	1183.33	1195.49
葡萄	764.91	777.07	789.23	801.39	813.55	825.71	837.87	850.03	862.19	874.35
柑橘类	412.95	425.11	437.27	449.43	461.59	473.75	485.91	498.07	510.23	522.39
梨	612.38	624.54	636.70	648.86	661.02	673.18	685.34	697.50	709.66	721.82
桃	10.00	22.16	34.32	46.48	58.64	70.80	82.96	95.12	107.28	119.44
猕猴桃	112.99	122.15	131.31	140.47	149.63	158.79	167.95	177.11	186.27	195.43
李子	1724.20	1736.36	1748.52	1760.68	1772.84	1785.00	1797.16	1809.32	1821.48	1833.64
西瓜	1839.14	1851.30	1863.46	1875.62	1887.78	1899.94	1912.10	1924.26	1936.42	1948.58

表 21 蔬菜种植面积（单位：千公顷）

	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	2021 年	2022 年	2023 年	2024 年
土豆	5368.53	5392.65	5416.77	5430.89	5455.01	5467.13	5491.25	5515.37	5519.49	5525.61
西红柿	1012.44	1036.56	1060.68	1070.80	1094.92	1119.04	1143.16	1155.28	1179.40	1191.52
大白菜	914.68	938.80	962.92	987.04	1011.16	1025.28	1049.40	1061.52	1056.64	1072.76
茄子	781.65	795.77	809.89	824.01	838.13	852.25	866.37	880.49	894.61	908.73
菠菜	640.50	664.62	688.74	712.86	736.98	761.10	785.22	809.34	833.46	857.58
胡萝卜	456.91	470.03	483.15	496.27	509.39	522.51	535.63	548.75	561.87	574.99
青椒	712.09	736.21	760.33	784.45	808.57	832.69	856.81	880.93	905.05	929.17
菜花	462.78	471.90	481.02	490.14	499.26	508.38	517.50	526.62	535.74	544.86
蘑菇	19.04	23.16	27.28	31.40	35.52	39.64	43.76	47.88	52.00	56.12

由表 14-15 可以看出水果和蔬菜的产量从战略上来说应该在逐年增加，而由表 16-19 可以看出：虽然果蔬的进出口量每年都有波动，但是相比总产量可以忽略不计。所以总的来说我国对果蔬发展战略是产量逐年增加的。而从表 20-21 可以看出各种果蔬的种植面积也应相应缓慢增长，所以说在未来我国人均对果蔬的摄入量也应逐渐增加。

## 五 政策建议

通过预测结果可以看出，我国居民的果蔬消费、营养健康，与合理膳食营养模式都有所偏离，相关部门需要进行合理的消费干预、采取相应的政策措施，从人体营养健康角度对我国果蔬的种植与消费进行宏观调控。

### （1）优化产业结构，完善食品体系

由于我国蔬菜和水果品种繁多，地域分布较广，人们能购买到的主要水果和蔬菜大概只有十种左右，所以这些主要的水果、蔬菜品种不仅总计产量应分别超过它们各自总产量的 90%，而且品种所蕴含营养素在成分和含量上都要满足人们的需要，能够保障人体健康所需的营养。根据问题四的结果在今后的产量规模的战略调整是：适当减少苹果、柑橘类、梨、西瓜、土豆、西红柿和茄子的种植面积，加大对桃、香蕉、葡萄、猕猴桃、李子、菠菜、胡萝卜、青椒、菜花的种植。只有这样才能补充缺乏的营养素，调整过剩的营养素。从而我们要树立“营养指导消费，消费引导生产”的观念，坚持“生产结构依据消费需求，消费需求符合营养健康”的原则，建立以营养健康为主要目标的果蔬产业结构调整模式。在保持品种多样化、区域化、优势化、专业化的形势下，变“种什么吃什么”为“吃什么种什么”，了解不同地区、不同人群的营养需求，调整当前蔬菜水果产业结构与层次，根据当地自然条件，合理规划种植面积，提高营养丰富的果蔬产量，保证营养均衡。

考虑居民的购买成本和种植者的收益以及人体摄入的营养均衡，同时兼顾果蔬产品的进出口贸易量和土地面积等因素，以及根据未来主要水果和蔬菜的人均消费量，得出应适当减少苹果、柑橘类、梨、西瓜、土豆、西红柿和茄子等的种植面积，加大对桃、香蕉、葡萄、猕猴桃、李子、菠菜、胡萝卜、青椒、菜花的种植。

### （2）加强宏观指导，制定相关法规

在今后相当长的一段时间里，政府及相关部门应掌握好我国膳食营养结构转型的关键时期，加强引导食品消费。抓紧进行食物与营养相关法律法规的研究工作，适时开展营养改善条例的立法工作，制定和实施食物消费与营养发展相关法规政策，将国民营养与健康改善工作纳入国家规划；各地区人民政府要根据相关纲要确立的目标、任务和重点，结合本地区实际，制定当地食物与营养发展实施计划。针对食物与营养的突出问题，依法规范食物生产经营活动，开展专项治理整顿，营造安全、诚信、公平的市场环境。创新食物与营养执法监督，提高行政监管效能。弘扬勤俭节约的传统美德，形成厉行节约、反对浪费的良好社会风尚。

### （3）宣传普及知识，倡导合理膳食

相关部门应加强社会公众的宣传教育，利用多种新闻媒体采取多种形式广泛宣传，增加宣传频率，加强对居民食物与营养的指导，引导居民改善食物与营养状况。提高全民营养意识，提倡健康生活方式，树立科学饮食理念，引导全民建立合理的食物消费模式和科学的膳食模式。相关部门应开展全国居民营养与基本健康监测工作，进行食物消费调查，定期发布中国居民食物消费与营养健康状况报告，向社会公开食物营养与健康调查数据库，实现信息共享。并针对不同区域地区居民食物消费的特点，发布适宜不同人群特点的膳食指南，增强营养知识传播的科学性。

#### (4) 调整收入格局，平衡地区发展

经济发展制约着食品消费，我国应继续促进经济的发展，提高人民收入水平和生活条件，完善社会保障体系和社会保障资金筹措运营机制，稳定居民的支出预期；发挥政府再分配调节能力，缩小居民收入差距，使中低、低收入者的食物结构得到改善。国家应从全国食物发展的全局和国民经济总体布局上加以考虑，充分考虑不同地域的特点和差异性，制定科学的食物发展战略：实行区域之间的互补互利，加强区域之间的合作和食物资源利用。区域根据自身发展条件，坚持发展区域特色，种植符合当地自然条件的优质果蔬，并进一步促进区域间的食物区际交流，提高蔬菜水果的进出口数量，确保地区居民的膳食均衡。加强农村商贸与流通基础设施建设，将城镇现代流通业向广大农村地区延伸，推进“万村千乡”市场工程，开拓农村食物市场，方便农村居民购买食物。

#### (5) 加快科技创新，重视特殊人群

针对食物、营养和健康领域的重大需求，引导企业加大食物与营养科技投入，加强对食物与营养重点领域和关键环节的研究；加强对新食物资源开发和食物安全风险分析技术的研究，在科技创新中提高食物安全水平；加强食物安全监测预警技术研究，促进食物安全信息监测预警系统建设。深入研究食物、营养和健康的关系，及时修订居民膳食营养素参考摄入量标准。针对幼、孕、老等特殊人群的营养问题，国家需高度重视。做好农村儿童营养保障工作，降低青少年生长迟缓、缺铁性贫血的发生率；遏制城镇儿童青少年超重、肥胖增长态势。做好孕产妇营养均衡调配，重点改善低收入人群孕妇膳食中钙、铁、锌和维生素 A 摄入不足的状况，预防中高收入人群孕妇因膳食不合理而导致的肥胖、巨大儿等营养性疾病。研究开发适合老年人身体健康需要的食物产品，重点发展营养强化食品和低盐、低脂食物。开展老年人营养监测与膳食引导，科学指导老年人补充营养、合理饮食，提高老年人生活质量和健康水平。

## 六 模型的灵敏度分析

### 6.5.1. 参数率的灵敏度分析

考虑 COWA 算子中参数  $\lambda$  对求解结果的影响，结果见表 22 所示。

表 22  $\lambda$  对求解结果的影响

$\lambda$ 变动	水果人均消费量 变动范围	蔬菜人均消费量 变动范围
2%	0.0487%	0.0231%
5%	0.0938%	0.0874%
10%	0.1043%	0.1134%

由表 22 可以看出，随着参数  $\lambda$  变动范围的扩大，人均消费量的变动范围也相应变大，但是整体上来说变动幅度都不是特别大。

### 6.2 果蔬多目标规划模型中参数 $\varphi$ 的灵敏度分析



参数 $\varphi$ 的变动对最终的果蔬消费量会产生影响。这里考虑 $\varphi$ 变动至 10%时果蔬消费量的变动情况，结果见表 23 所示。

表 23 参数 $\varphi$ 对果蔬消费量的影响

$\varphi$ 变动	水果人均消费量 变动范围	蔬菜人均消费量 变动范围
2%	0.0123%	0.0451%
5%	0.0342%	0.1132%
10%	0.1533%	0.1440%

由表 23 可以看出，随着参数 $\varphi$ 变动范围的扩大，果蔬人均消费量变动范围也相应变大，但变动幅度也很小。

## 七 模型的优缺点分析和改进

### 7.1 模型优点

- (1) 采用组合预测方法对各种果蔬的消费量进行预测。因为需要做预测的果蔬个数很多，每个单项预测方法可能适用这个水果的预测，而不适合另个蔬菜的预测。采用组合预测方法就可以避免这个问题，提高准确率；
- (2) 大量使用启发式算法，提高计算速度与精度；
- (3) 多次对数据进行区间处理、模糊处理，更有效的解决了本文一些不确定数据难计算问题；
- (4) 多次采用最优化模型，行之有效的解决各种果蔬消费量大小问题，即对各种果蔬消费的分配问题；

### 7.2 模型缺点

- (1) 在得出结果的过程中需要产生大量随机数，故结果自身存在波动性。需要多次运行程序以提高准确度。
- (2) 由于建模前确定部分假设，故误差难以避免。

### 7.3 模型的进一步改进

- (1) 从模型角度，可以通过尝试其他算法模型，从而寻找最佳模型，提高程序的运行速度和模型的准确性，例如选择模拟退火、蚁群算法、神经网络算法等智能算法。
- (2) 设定更多果蔬生产方面因素，应用上述模型进行求解，进一步提高适用性。

## 八 参考文献

- [1]林海明.因子分析精确模型的基本思想与方法[J].统计与信息论坛, 2006, 5:23-25.
- [2]方述诚,汪定伟.模糊数学与模糊优化[M].北京:科学出版社,1997, 34-44.
- [3] Kennedy J, Eberhart R C. Particle swarm optimization[C]. Proc. IEEE Int. Conf.on Neural Networks, Perth, WA,Australia, 1995. 1942-1948
- [4] Eberhart R C, Kennedy J A. A new optimizer using particle swarm theory[C]. Proc. The Sixth Int. Symposium on Micro Machine and Human Science, Nagoya, Japan, 1995. 39-43
- [5] Kennedy J, Eberhart R C, Shi Y. Swarm intelligence[M]. San Francisco: Morgan Kaufmann Publishers, 2001
- [6] Dautenhahn K. Book review: Swarm intelligence[J]. Genetic Programming and Evolvable Machines, 2002,3(1):93-97
- [7]邓聚龙.灰色预测与决策[M].武汉:华中理工大学出版社,1988.
- [8]刘思峰,郭天榜,党耀国.灰色系统理论及应用[M].北京:科学出版社,2000.
- [9]熊岗,陈章潮.灰色预测模型的缺陷及改进方法[J].系统工程,1992,10(2):42-44.
- [10]吉培荣,黄巍松,胡翔勇.无偏灰色预测模型[J].系统工程与电子技术,2000,22(6):6-8.
- [11]朱孔来.灰色马尔柯夫链预测模型及其应用[J].系统工程理论与实践,1993, (2) :33 - 37.
- [12]吴今培.实用时序分析[M].长沙:湖南科学技术出版社,1989
- [13]吴今培,段方勇.基于神经网络的非线性时间序列预测方法研究[J].系统工程,1997,15(5):61-64
- [14]张立明.人工神经网络的模型及其应用[M].上海:复旦大学出版社,1993.
- [15]向国全,董道珍.BP 模型中的激励函数和改进的网络训练法[J].计算机研究与发展,1997,34(2):113-117.
- [16]李学伟,关忠良,陈景艳.经济数据分析预测学[M].北京:中国铁道出版社,1998.
- [17]Box G E P, Jenkins G M. Time series analysis, forecasting and control[M]. San Francisco, Holden Day, 1970.
- [18]曹星平, 易东云, 吴翊等. 基于神经网络的时间序列预测方法进展[J]. 电脑与信息技术, 1999, (6): 1-3.
- [19] Bates J M, Granger C W J. Combination of forecasts [J]. Operations Research Quarterly, 1969, 20(4) : 451 ) 468.
- [20]唐小我.组合预测计算方法研究[J].预测,1991(4):35-39.
- [21]唐小我.最优组合预测方法及其应用[J].数理统计与管理,1992,11(1):31-35.
- [22]陈华友.组合预测方法有效性理论及其应用[M].北京:科学出版社,2008:56-98
- [23]沈家骅,严振祥.基于区间分析的组合预测系数确定方法[J].武汉理工大学学报,2006, 30(6):1077-1080.
- [24]陈华友,刘春林.基于 IOWA 算子组合预测方法[J].预测,2003,22(6):61-65.
- [25]王晓,刘兮,陈华友,周礼刚.基于 IOWA 算子的区间组合预测方法[J].武汉理工大学学报:信息与管理工程版,2010,32(2):221-225.
- [26]Hayes B. A lucid interval[J].American Scientist,2003(91):484-488.

- [27]Yager R.R.The power average operator[J]. IEEE Transactions on Systems, Man and Cybernetics A,2001, 31:724-731.
- [28]金飞飞,李捷,陈华友,周礼刚.基于COWA算子的区间组合预测方法[J].武汉理工大学学报: 信息与管理工程版,2013,2095-3852(2013)05-0668-05.
- [29]蔡中华等.应用自组织映射对中国地区产业结构的聚类分析, 复杂系统与复杂性研究, 2004.
- [30]郝俊卿, 王雁林.基于自组织映射神经网络模型的区域经济评价, 地域研究和开发, 2005.