



(12)发明专利申请

(10)申请公布号 CN 106295686 A
(43)申请公布日 2017.01.04

(21)申请号 201610624628.8

(22)申请日 2016.08.01

(71)申请人 安徽农业大学

地址 230036 安徽省合肥市蜀山区长江西路130号

(72)发明人 杨宝华 钱彬彬 戴前颖 谢申汝
徐光祥 王淑娟 杨玉洁

(74)专利代理机构 合肥市上嘉专利代理事务所
(普通合伙) 34125

代理人 王伟

(51)Int.Cl.

G06K 9/62(2006.01)

G06N 3/12(2006.01)

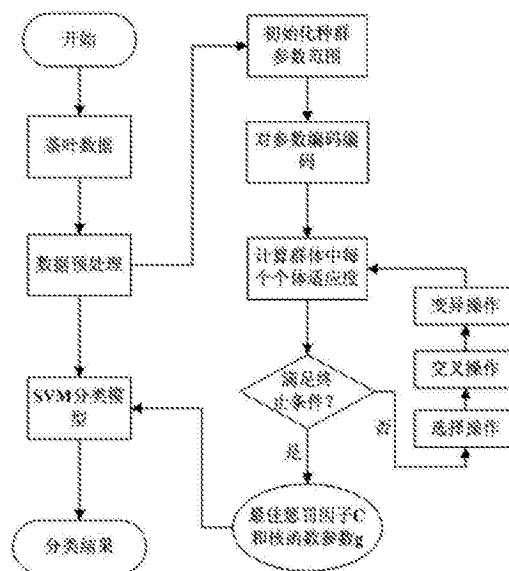
权利要求书2页 说明书8页 附图7页

(54)发明名称

参数优化支持向量机的茶叶香气分类方法

(57)摘要

本发明涉及一种参数优化支持向量机的茶叶香气分类方法,旨在通过改进的支持向量机解决茶叶香气分类问题,属于茶叶香气分类领域。其原理利用电子鼻传感器模拟人感官品评的功能和特征,采集不同时间不同传感器的特征值,构建数据集。利用算法优化支持向量机,获取优化的惩罚因子C和核函数参数g,进而构建毛峰茶叶香气的支持向量机SVM分类模型。本发明的有益效果在于可提高预测茶叶香气分类的效率和准确度,为消费者提供茶叶香气分类的有效方法。



1. 参数优化支持向量机的茶叶香气分类方法,其特征在於:通过电子鼻采集茶叶香气的特征数据作为输入样本,优化惩罚因子 C 和核函数参数 g ,构建茶叶香气的支持向量机SVM分类模型,并据此计算出茶叶香气分类。

2. 根据权利要求1所述参数优化支持向量机的茶叶香气分类方法,其特征在於:电子鼻采集的茶叶香气的特征数据包括:氢气,甲基烷烃,无机硫化物,羰基类和醇类,有机硫化物,长链烷烃指标值。

3. 根据权利要求2所述参数优化支持向量机的茶叶香气分类方法,其特征在於:电子鼻采集的茶叶香气的特征数据还包括:芳香苯,氮氧化合物,芳香氨类,芳香烷烃。

4. 根据权利1-3任一所述参数优化支持向量机的茶叶香气分类方法,其特征在於利用遗传算法优化支持向量机,获取优化的惩罚因子 C 和核函数参数 g 。

5. 根据权利1-3任一所述参数优化支持向量机的茶叶香气分类方法,其特征在於利用网格搜索算法优化支持向量机,获取优化的惩罚因子 C 和核函数参数 g 。

6. 根据权利要求4所述参数优化支持向量机的茶叶香气分类方法,其特征在於包括以下步骤:

步骤1:准备茶叶样品,利用PEN3型电子鼻采集茶叶香气的特征数据;

步骤2:构建样本集,对茶叶香气的特征数据进行归一化处理;

步骤3:设定支持向量机的输入向量数目、选定核函数以及参数的设定;

步骤4:利用遗传算法优化支持向量机,进行模型训练,输出优化后的支持向量机最佳惩罚因子 C 和核函数参数 g ;

步骤5:由电子鼻检测茶叶的测试集样本数据作为输入,通过步骤4优化后的支持向量机预测茶叶的香气分类。

7. 根据权利要求5所述参数优化支持向量机的茶叶香气分类方法,其特征在於包括以下步骤:

步骤1:准备茶叶样品,利用PEN3型电子鼻采集茶叶香气的特征数据;

步骤2:构建样本集,对茶叶香气的特征数据进行归一化处理;

步骤3:设定支持向量机的输入向量数目、选定核函数;

步骤4:利用网格搜索算法优化支持向量机,进行模型训练,输出优化后的支持向量机最佳惩罚因子 C 和核函数参数 g ;

步骤5:由电子鼻检测茶叶的测试集样本数据作为输入,通过步骤4优化后的支持向量机预测茶叶的香气分类。

8. 根据权利要求6参数优化支持向量机的茶叶香气分类方法,其特征在於,优化支持向量机最佳惩罚因子 C 和核函数参数 g 包括:

步骤4.1随机初始化支持向量机的 C 和 g ,对 C 和 g 进行编码;

步骤4.2构造个体适应度函数;

步骤4.3选择操作,使用选择操作处理个体适应度值,选择部分输出个体适应度值对应的个体;

步骤4.4交叉操作,使用交叉操作对个体进行处理,输出交叉操作结果;

步骤4.5变异操作,使用变异操作对交叉操作结果进行处理,得到新的基因值;

步骤4.6根据遗传算法结束条件,得到支持向量机最佳的惩罚因子 C 和核函数参数 g 。

9. 根据权利要求8参数优化支持向量机的茶叶香气分类方法,其特征在於,优化支持向量机最佳惩罚因子C和核函数参数g包括:

步骤4.1数据初始化:设定初始搜索步长、参数范围;

步骤4.2网格搜索建模方法:使用5折交叉验证方法进行模型构建;

步骤4.3根据网格搜索算法结束条件,得到支持向量机最佳的惩罚因子C和核函数参数g。

参数优化支持向量机的茶叶香气分类方法

技术领域

[0001] 本发明涉及茶叶香气分类方法,尤其是涉及一种参数优化支持向量机的茶叶香气分类方法。

背景技术

[0002] 气味是食品品质评价的一个重要指标,目前对食品气味的评判主要依靠有经验的专业人员来完成。主要采用感官定量描述分析法、主成分分析法和传统评分法相结合的方法来完成评估。同样的气味,不同的人很难得到一致的评价结果;甚至同一个人在不同的环境、不同的情绪时,对同一种气味也有不同的感受和评价。从而使得采用人的感官进行评判存在一定的局限性,为了减少评判的错误率,用机器模拟人的感官(如模拟人视觉的电子眼、模拟人的味觉的电子舌、模拟人的嗅觉的电子鼻等)进行食品品质评判已成为一种趋势。电子鼻作为一种新兴的感官分析技术,已经广泛应用于食品、烟草、医学、环境监测和爆炸物检测等领域。其中在新鲜度、成熟度判别、病害检测、品质监控、质量评价和安全检测中显示出独特优点,具有快速、无损、重复性好的优点。电子鼻主要由气味取样操作器、气味传感器阵列和信号处理系统组成,是利用气味传感器阵列的响应曲线来识别气味。由于电子鼻得到的是样品中挥发出的气味信息,而不是样品中某成分的定量或定性结果。茶叶的香气,主要是由芳香物质的种类、浓度决定的。电子鼻提供一种快速无损的方式来测定茶叶香气的挥发物质,从而对待测茶叶的气味进行客观评价。因此,电子鼻提供一种快速无损的方式来测定食品挥发物质,从而对待测样品的品质进行客观评价。

[0003] 茶叶的香气一般是由复混物组成,比较复杂,不同品种茶叶的香气不一样,甚至产地不同的同一种茶叶香气也不完全一样,尤其当茶叶贮放一段时间后,香气、滋味、颜色就会在短时期内发生变化。所以不同的人对同一种茶甚至同一杯茶香气有不同的看法,这可能是由个人的喜好、习惯、经验、敏感度造成的。因此,传统的香气判断方法无法正确快速的进行分类判断。而支持向量机算法(Support Vector Machine, SVM)是在统计学习理论上发展起来的一种分类方法,由Cortes和Vapnik在1995年首先提出的, SVM算法是根据结构风险最小化原则,在模型的复杂性和学习能力之间寻求最佳,以获得最优的泛化推广能力。在解决小样本问题、非线性问题以及高维数据等问题中表现出特有的优势,已被广泛应用于判断、处理预测、数据拟合、综合评价以及模式识别等诸多问题。

[0004] 在对茶叶气味分类的问题上,已经有了典型判别分析法及神经网络方法等,但是在面对大量茶叶香气数据时,这些算法都存在准确率不高的问题,本专利针对这个问题,采用算法优化SVM参数的方法,对茶叶的香气进行不同程度的分类,提高分类的准确率。

发明内容

[0005] 本发明要解决的技术问题是提供一种参数优化支持向量机的茶叶香气分类方法。本发明是通过以下技术方案来实现的。

[0006] 通过电子鼻采集茶叶香气数据,将检测到的芳香苯,氮氧化合物,芳香氨类,氢气,

芳香烷烃,甲基烷烃,无机硫化物,羰基类和醇类,有机硫化物,长链烷烃指标的校正集数据作为输入样本,通过算法优化惩罚因子C和核函数参数g,构建支持向量机进行自分类。得到训练好的茶叶分类模型,并据此计算出茶叶预测集数据的分类准确率。惩罚因子C和核函数参数g通过遗传算法或网格搜索法优化得到。

[0007] 本发明的有益效果:

[0008] 本发明将遗传算法优化支持向量机和网格搜索法优化支持向量机应用于实际的茶叶数据中,对茶叶香气进行分类,有效地提高了茶叶香气的分类准确率,保证了茶叶的鲜嫩口感。可提高预测茶叶香气分类的效率和准确度,为消费者提供茶叶香气分类的有效方法。

附图说明

- [0009] 图1为毛峰茶叶样本集主成分分析的载荷图。
- [0010] 图2为本发明遗传算法优化支持向量机毛峰茶叶香气分类方法的流程示意图。
- [0011] 图3为毛峰茶叶香气的GA-SVM算法运行10次的准确率。
- [0012] 图4为毛峰茶叶香气的GA-SVM算法运行10次的建模时间。
- [0013] 图5为毛峰茶叶香气的遗传算法在训练过程中的适应度曲线。
- [0014] 图6为GA-SVM算法毛峰茶叶香气可视化结果。
- [0015] 图7为SVM算法毛峰茶叶香气可视化结果。
- [0016] 图8为网格搜索算法优化支持向量机毛峰茶叶香气分类方法的流程示意图。
- [0017] 图9为毛峰茶叶香气分类的SVM参数选择等高线图。
- [0018] 图10为毛峰茶叶香气分类的SVM参数选择3D视图。
- [0019] 图11为GridSearch-SVM算法的毛峰茶叶香气分类可视化结果图。
- [0020] 图12为SVM算法的毛峰茶叶香气分类可视化结果图。

具体实施方式

[0021] 实施例1:

[0022] 下面根据附图和实施例对本发明作进一步详细说明,以毛峰茶叶为例。

[0023] 步骤1:确定毛峰茶叶检测指标;

[0024] 步骤1.1毛峰茶叶样品的准备;

[0025] 各个等级的毛峰茶叶分别准备70个重复样品,每个重复样品质量为5g,双层薄膜密封在500ml的烧杯中,室温下静置45min。

[0026] 根据感官评审时的要求,茶、水比1:50,取5g茶叶用250ml水冲泡。泡茶用水为沸滚适度100℃的纯净水,冲泡时间为5min,然后将茶水滤出。将茶水和茶底分别在500ml的烧杯中密封、静置45min使得烧杯顶空富集茶叶挥发性成分的同时,水温也冷却至室温,室温保持 $25 \pm 3^\circ\text{C}$ 。由于茶水和茶底中水蒸气很多,所以检测时要放些硅胶,以减少水蒸气的影响。同样对每个等级的茶叶也各准备70个重复。

[0027] 由于这部分的样品中水蒸气含量较多,虽然用硅胶预处理过,但水蒸气含量仍相对较大,所有要结合吸附/解吸附单元使用,对检测的挥发性成分进行富集、浓缩和清洁,以减少对传感器阵列的影响。把吸附/解吸附单元与电子鼻系统连接好,运行其配套气味指纹

分析处理软件,首先检测吸附/解吸附单元是否连接好;然后设定系统各参数,即可开始检测。

[0028] 步骤1.2毛峰茶叶指标检测;

[0029] 电子鼻采用德国Airsense公司生产的PEN3型电子鼻,含有10个不同的金属氧化物传感器。电子鼻的原理是利用特定的金属氧化物和生物膜,根据挥发性物质分子接触引起膜电位微小的变化来判断是否有气味和气味的强弱,以特定的传感器和模式识别系统进行快速提供被测样品的全面信息,提示样品的隐含特征。传感器的输出为传感器接触挥发性气体的电导率与传感器经过活性炭过滤后纯净空气的电导率之比。得到10个传感器特征值。这种气敏传感器具有可靠性高、灵敏度好和重复性强等特点。

[0030] 电子鼻PEN3的标准传感器阵列见表1。

[0031] 表1PEN3的标准传感器阵列

阵列序号	传感器名称	传感器表示符号	性能描述
1	W1C	R(1)	芳香苯
2	W5S	R(2)	氮氧化合物
3	W3C	R(3)	芳香氨类
4	W6S	R(4)	氢气
[0032] 5	W5C	R(5)	芳香烷烃
6	W1S	R(6)	甲基烷烃
7	W1W	R(7)	无机硫化物
8	W2S	R(8)	羰基类和醇类
9	W2W	R(9)	有机硫化物
10	W3S	R(10)	长链烷烃

[0033] 步骤2毛峰茶叶建立样本集;

[0034] 步骤2.1毛峰茶叶样本数据获取;

[0035] 由电子鼻检测到毛峰茶数据的部分数据的详细信息情况见表2。其中,1到10号传感器记录在不同时间下,传感器接触挥发性气体的电导率与传感器经过活性炭过滤后纯净空气的电导率之比,香气是感官审评的陈化程度值。该数据集包括452个样本,10个样本属性,2个类别。

[0036] 表2毛峰茶数据情况

[0037]

1号传感器	2号传感器	3号传感器	4号传感器	5号传感器	6号传感器	7号传感器	8号传感器	9号传感器	10号传感器	香气
0.8454	3.776	0.9052	1.0455	0.9554	2.1057	17.9098	1.3894	6.3965	1.2965	0
0.835	3.7834	0.9006	1.0462	0.9546	2.154	18.0484	1.4107	6.386	1.3076	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

[0038]

0.8422	3.5353	0.9035	1.0369	0.9623	1.9677	15.5569	1.3076	6.0947	1.2018	0
0.8305	3.544	0.8984	1.0373	0.9605	2.0114	15.6912	1.3234	6.0844	1.2089	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.6139	3.6806	0.7994	1.0845	0.9301	2.9601	20.5903	1.8495	6.5863	1.5192	1
0.6113	3.6673	0.7998	1.0846	0.9302	2.9667	20.5053	1.8584	6.5535	1.5196	1

[0039] 步骤2.2毛峰茶叶样本数据主成分分析

[0040] 对该样本集进行主成分分析,得到6个主成分,其载荷图如图1所示,提取的6个主成分分别为:氢气,甲基烷烃,无机硫化物,羰基类和醇类,有机硫化物,长链烷烃指标值。

[0041] 步骤2.3毛峰茶叶样本归一化处理;

[0042] 为了使分类更准确,对毛峰茶叶样本数据进行归一化处理。使所有的样本数据都归一为[0,1]之间,对于输出向量通过反归一化处理,得到原值。采用的数据归一化处理计算公式为:

[0043]
$$P' = (P - P_{\min}) / (P_{\max} - P_{\min})$$

[0044] 式中, P' 、 P 、 P_{\max} 、 P_{\min} 分别为归一化后样本数据,原始样本数据,原始样本数据的最大值和最小值。

[0045] 步骤3支持向量机的输入向量数目,选定核函数;

[0046] 步骤3.1输入向量数目,选定核函数;

[0047] 对毛峰茶叶样本数据按照校正集和测试集4:1进行划分,校正集作为支持向量机的输入向量数目,支持向量机核函数选用径向基(RBF)核函数。

[0048] 步骤3.2参数的设定;

[0049] 支持向量机惩罚因子C的范围设为[0,100],核函数参数g的范围设为[0,1000]。

[0050] 遗传算法种群规模的大小关系着算法性能的好坏。种群规模太大,计算时间就会增加,同时降低了算法的竞争力。种群规模太小,就会导致算法优化的质量不高。本专利选取种群规模选取为20。

[0051] 遗传算法迭代次数是表示遗传算法运行结束条件的一个参数,本实施例迭代次数选取为200。

[0052] 遗传算法交叉概率一般取值较大,但如果太大,它会破坏群体中的优良模式,对进化运算不利。本实施例交叉概率选取为0.7。

[0053] 遗传算法变异概率取值太大,则容易破坏群体中的优良模式,使得遗传算法的搜索趋于随机性;若取值过小,则它产生新个体和抑制早熟的能力会较差。本实施例变异概率选取为0.1。

[0054] 步骤4利用遗传算法优化支持向量机,以下又称GA-SVM,进行模型训练,输出最佳惩罚因子C和核函数参数g;

[0055] 步骤4.1模型训练;

[0056] 本实施例采用遗传算法优化支持向量机对毛峰茶香气品质进行分类。遗传算法是基于自然选择的生物进化,是一种模拟生物进化过程的随机方法。是一种具有自适应能力的、全局性的概率搜索算法。

[0057] 步骤4.2使用遗传算法优化支持向量机惩罚因子C和核函数参数g。具体包括:

[0058] 步骤4.2.1编码工作;

[0059] 对数值进行二进制编码来离散化,使用的编码符号集由二进制符号0和1所组成的二值符号集 $\{0,1\}$,它所构成的个体基因型是一个二进制编码符号串。假定其进化次数为 $t=1, \dots, T$;假定长度为 n 的染色体表示为符号串 $x = x_1, x_2, \dots, x_n$,其中:记号 $x_i (i=1, 2, \dots, n)$ 代表一个遗传基因,所有等位基因的组合构成了解的基本空间:

$$[0060] \quad A = x_1 \times x_2 \times \dots \times x_n = \prod_{i=1}^n x_i$$

[0061] 步骤4.2.2构造个体适应度函数;

[0062] 遗传算法常将目标函数直接作为适应度函数,但由于在执行选择操作时,它要按与个体适应度成正比的概率来决定当前群体中每个个体遗传到下一代群体中的几率,而要正确计算此概率,要求所有个体的适应度值必须非负所以实际操作时。将待求解的目标函数做适当处理后再转化为适应度函数,即:

$$[0063] \quad \text{Fit}(f(x)) = \begin{cases} c_{\max} - f(x), & f(x) < c_{\max} \\ 0 & \text{其他} \end{cases}$$

[0064] 式中 c_{\max} 为发 $f(x)$ 的最大估计值。

[0065] 步骤4.2.3选择操作;

[0066] 根据适者生存的原则,选择适应度高的个体,可有效地避免有效基因的损失,使高性能的个体得以较大的概率生存,从而提高全局收敛性和计算效率。

[0067] 选择操作采用轮盘赌操作,令 $pp_i = \sum_{j=1}^{NP} p_j$, $pp_0=0$,计算公式为:

$$[0068] \quad p_i = \frac{\text{fitness}(x_i)}{\sum_{i=1}^{NP} \text{fitness}(x_i)}$$

[0069] pp_i 为累计概率, p_i 为个体的选择概率, $\text{fitness}(X_i)$ 为个体适应度。共转轮 NP 次(NP 为种群个体数),每次转轮时,随机产生0到1之间的随机数 r ,当 $pp_{i-1} \leq r \leq pp_i$ 时选择个体 i 。

[0070] 步骤4.2.4交叉操作;

[0071] 交叉选择是将两个选择出来的个体(染色体)在交叉点进行交叉互换,重组后生成两个崭新的子代个体(染色体),这两个新的子代个体的性状是由它们父代个体的性状组合而成的。假设在两个个体 X_A^i 、 X_B^i 之间进行算术交叉、则交叉运算后所产生出的两个新个体是:式中, α 为一参数。

$$[0072] \quad \begin{cases} X_A^{i+1} = \alpha X_B^i + (1 - \alpha) X_A^i \\ X_B^{i+1} = \alpha X_A^i + (1 - \alpha) X_B^i \end{cases}$$

[0073] 步骤4.2.5变异操作;

[0074] 变异操作就是由于一些偶然因素,引起了基因结构的改变,从而产生了新的染色体。

[0075] 变异操作的一般过程是:

[0076] 1.从种群中任意的选取某个个体(染色体)。

[0077] 2.以某个概率对该个体的染色体编码的某一个位置上的字符进行改变。得到了变

异后的个体(染色体)。

[0078] 使用变异操作对交叉操作结果进行处理,得到新的基因值 x'_k 由下式确定:

$$[0079] \quad x'_k = \begin{cases} x_k + \Delta(t, U_{\max}^k - v_k), & \text{random}(0,1) = 0 \\ x_k - \Delta(t, U_{\max}^k - v_k), & \text{random}(0,1) = 1 \end{cases}$$

[0080] 式中,random(0,1)表示随机取0或1, v_k 为 x_k 取值范围中的一个随机数, $\Delta(t,y)$ (y 代表 $U_{\max}^k - v_k$ 和 $U_{\max}^k + v_k$)表示 $[0,y]$ 范围内符合非均匀分布的一个随机数。

[0081] 步骤4.2.6得到最佳的惩罚因子C和核函数参数g;

[0082] 当达到最大迭代次数,通过解码得到的最优惩罚因子C和核函数参数g,此时,C=0.515,g=6.998,遗传算法优化支持向量机毛峰茶叶香气的流程图如图2所示。

[0083] 步骤5毛峰茶叶的香气分类;

[0084] 根据上述步骤,本实施例选用支持向量机和遗传算法优化支持向量机对毛峰茶叶进行香气分类,由于优化的结果具有一定的随机性,对遗传算法优化支持向量机算法进行10组运算,10次运算所得C、g和时间t如表3,准确率如图3所示。建模时间如图4所示。对其运算指标取均值进行对比分析。计算结果如表4所示。遗传算法在训练过程中的适应度曲线如图5所示,当遗传算法运行到110代以后平均适应度值已增加到最大值且趋于平稳状态,而最佳适应度值随着迭代次数的增加一直保持不变,直到运行到200代后,算法停止。此时,输出的SVM最优参数C=0.665,g=7.843,建模时间为54.98s。

[0085] 表3遗传算法优化支持向量机10次C和g值

[0086]

次数	1	2	3	4	5	6	7	8	9	10
C	0.515	1.614	0.388	2.619	0.837	0.408	0.441	0.665	0.757	1.904
g	7	9.156	7.957	6.756	7.887	7.713	6.82	7.843	7.777	7.09
t	52.3	49.99	57.88	50.28	43.94	54.51	49.44	54.98	55.22	56.77

[0087] 表4显示了采用支持向量机算法和遗传算法优化支持向量机算法对毛峰茶叶的香气分类对比结果,从表中可以看出,在建模时间上,支持向量机具有相对更快的建模速度,只用了0.25s;在准确率上,遗传算法优化支持向量机的准确率相对高很多,与支持向量机相比,高了33.33%。结果表明,遗传算法优化支持向量机对茶叶的香气的分类准确率更高。尤其是在面对要求高精度时,效果更显著。

[0088] 表4分类结果对比

	算法	建模时间/s	分类准确率/%
[0089]	SVM	0.25	59.05
	GA-SVM	54.98	92.38

[0090] 茶叶数据集包含2个类别属性:“0”、“1”。GA-SVM算法的毛峰茶叶香气可视化结果如图6所示。SVM算法的毛峰茶叶香气可视化结果如图7所示。图中X轴表示测试集样本,Y轴表示类别标签,图中圆形表示实际测试集分类,星形表示预测测试集分类。

[0091] 本实施例利用遗传算法的计算时间少,鲁棒性高、具有很好的收敛性等优点,将该

方法应用于支持向量机算法的惩罚因子C和核函数参数g的选取,进而构建GA-SVM毛峰茶叶香气分类模型,本发明不仅为消费者提供一种有效的茶叶香气分类方法,而且进一步提高了分类精度,得到较好的结果。

[0092] 实施例2

[0093] 使用网格搜索算法优化支持向量机,以下又称GridSearch-SVM,得出最佳惩罚因子C和核函数参数g。将实施例1中的步骤4和步骤5替换为以下技术方案:

[0094] 步骤4利用网格搜索算法优化支持向量机,进行模型训练,输出最佳惩罚因子C和核函数参数g;

[0095] 步骤4.1模型训练;

[0096] 本实施例采用网格搜索算法优化支持向量机对毛峰茶香气品质进行分类。针对本发明的支持向量机而言,所需要搜索的参数组为惩罚因子C和核函数参数g。因此,可以将这两个参数在二维空间划分一个网格并通过遍历测试的方式寻找最优参数组。首先,给出这两个参数的搜索范围并设定其移动的步距,通常为0.1。之后根据所设置的移动步距逐一取值测试。对于所选取的惩罚因子C和核函数参数g利用5阶交叉验证方法得到其在该组参数下校正集的分类准确率,使校正集分类准确率最高的那组C和g就是所要寻找的最优参数。网格搜索算法优化支持向量机对毛峰茶叶分类的流程图如图8所示。

[0097] 步骤4.2使用网格搜索算法优化支持向量机惩罚因子C和核函数参数g;

[0098] 步骤4.2.1数据初始化:设定初始搜索步长、参数范围;

[0099] 本实施例中选择参数C的搜索范围围为 $[-8, 8]$,搜索步长为0.8,参数g的搜索范围为 $[-8, 8]$,搜索步长为0.8。

[0100] 步骤4.2.2网格搜索建模方法;

[0101] 根据茶叶预处理后的数据,使用5折交叉验证方法(将校正集分为互不交叉的5个子集,轮流选择其中的4个子集进行参数训练,剩余的一个子集对选择的参数进行验证)计算出每组惩罚因子C和核函数参数g分类准确率。

[0102] 步骤4.2.3根据网格搜索算法结束条件,得到支持向量机最佳的惩罚因子C和核函数参数g。其中: $C=0.25, g=2.83$ 。

[0103] 步骤5毛峰茶叶的香气分类;

[0104] 根据上述方法,本实施例选用支持向量机和网格搜索法优化支持向量机参数对毛峰茶叶进行分类,对其运算指标取均值进行对比分析。计算结果如表3所示。网格搜索法参数选择结果如图9、图10所示。从图可以看出,取 $-2 < \log_2 c < 0, -2 < \log_2 g < 2$ 时,此时分类效果最好,此时,SVM最好,此时,SVM最优参数 $C=0.25, g=2.83$,建模时间为5.69s。

[0105] 表5分类结果对比

	算法	建模时间/s	分类准确率/%
[0106]	SVM	0.25	59.05
	GridSearch-SVM	5.69	91.74

[0107] 表5显示了采用支持向量机算法和网格搜索法优化支持向量机参数对毛峰茶叶的分类性能对比结果,从表中可以看出,在建模时间上,支持向量机算法具有相对更快的建模速度,只用了0.25s;在分类准确率上,网格搜索优化支持向量机的分类准确率相对高很多,

与支持向量机相比,高了31.69%,其准确率与基于遗传算法优化的支持向量机分类相当,但是建模时间远远小于GA-SVM。结果表明,网格搜索优化支持向量机的具有更高的分类准确率。网格搜索优化支持向量机所得到的分类模型分类效果更好,尤其是在面对大数据时,效果更显著。

[0108] 茶叶数据集包含2个类别属性:“0”、“1”。GridSearch-SVM算法的毛峰茶叶香气可视化结果如图11所示。SVM算法的毛峰茶叶香气可视化结果如图12所示。图中X轴表示测试集样本,Y轴表示类别标签,图中圆形表示实际测试集分类,星形表示预测测试集分类。

[0109] 本实施例利用网格搜索法在参数空间上取若干分格,遍历输入空间中所有网格交叉点,得到最优解。将该方法应用于支持向量机算法的惩罚因子C和核函数参数g的选取,进而构建GridSearch-SVM毛峰茶叶香气分类模型,本实施例不仅为消费者提供一种有效的茶叶香气分类方法,而且进一步提高了分类精度,得到较好的结果。

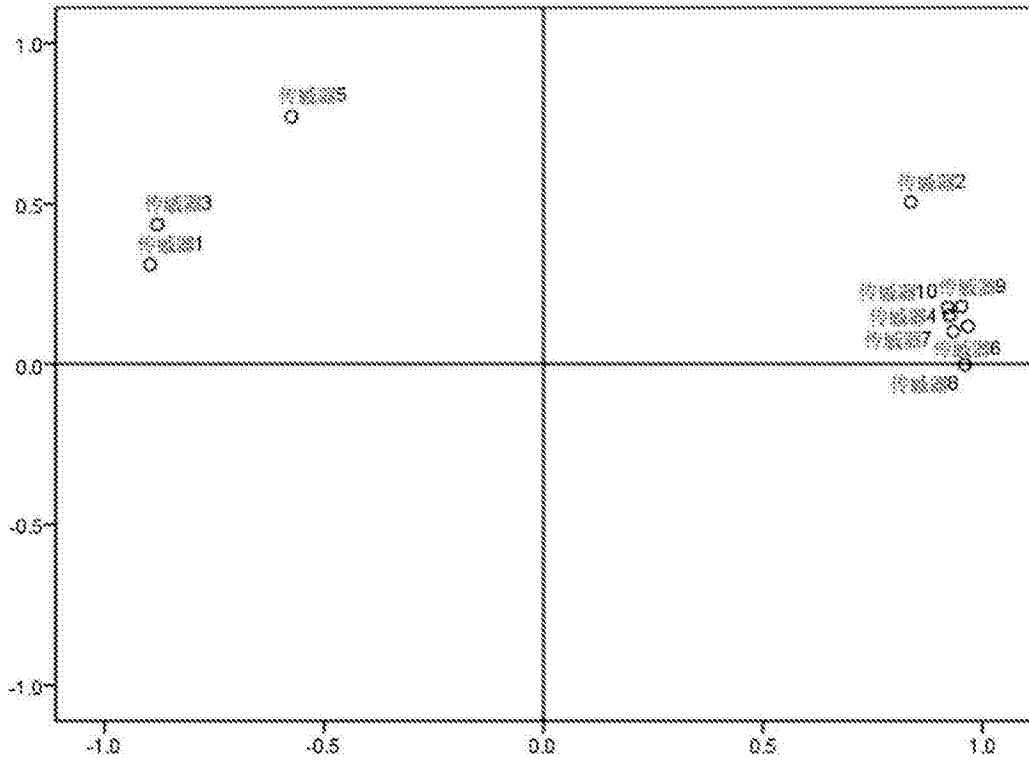


图1

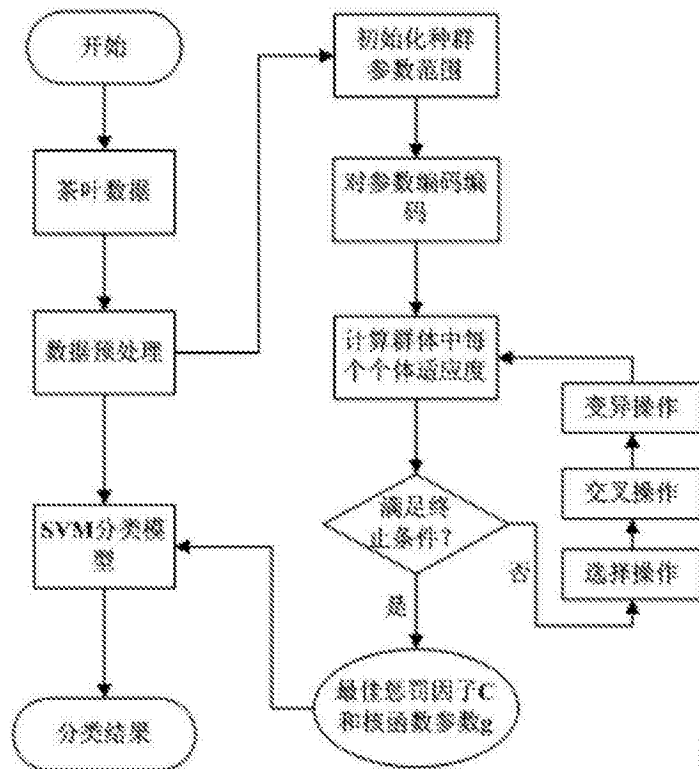


图2

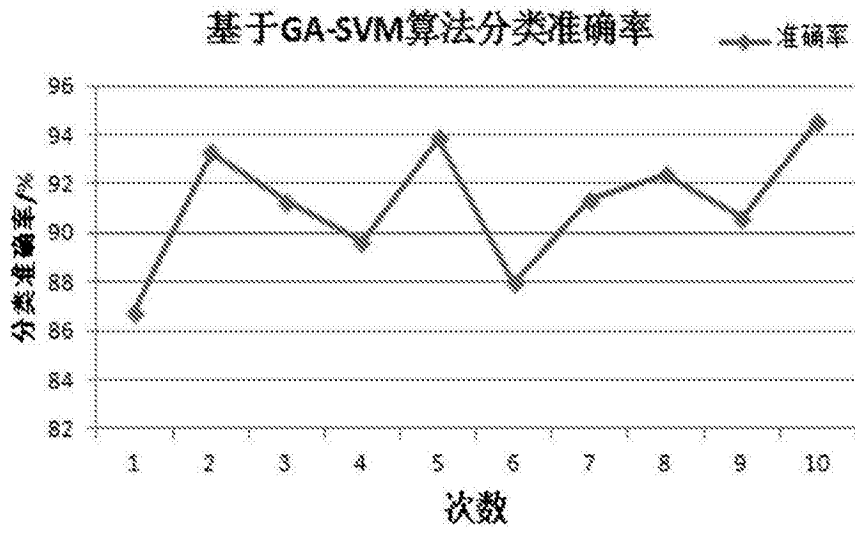


图3

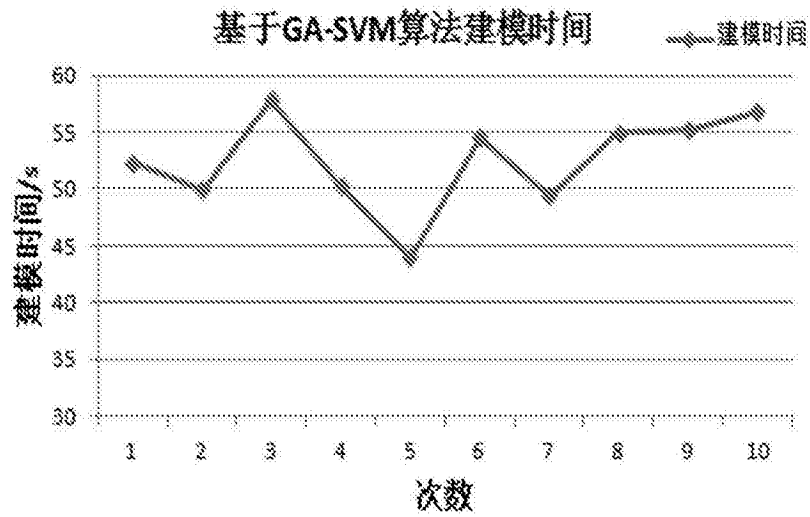


图4

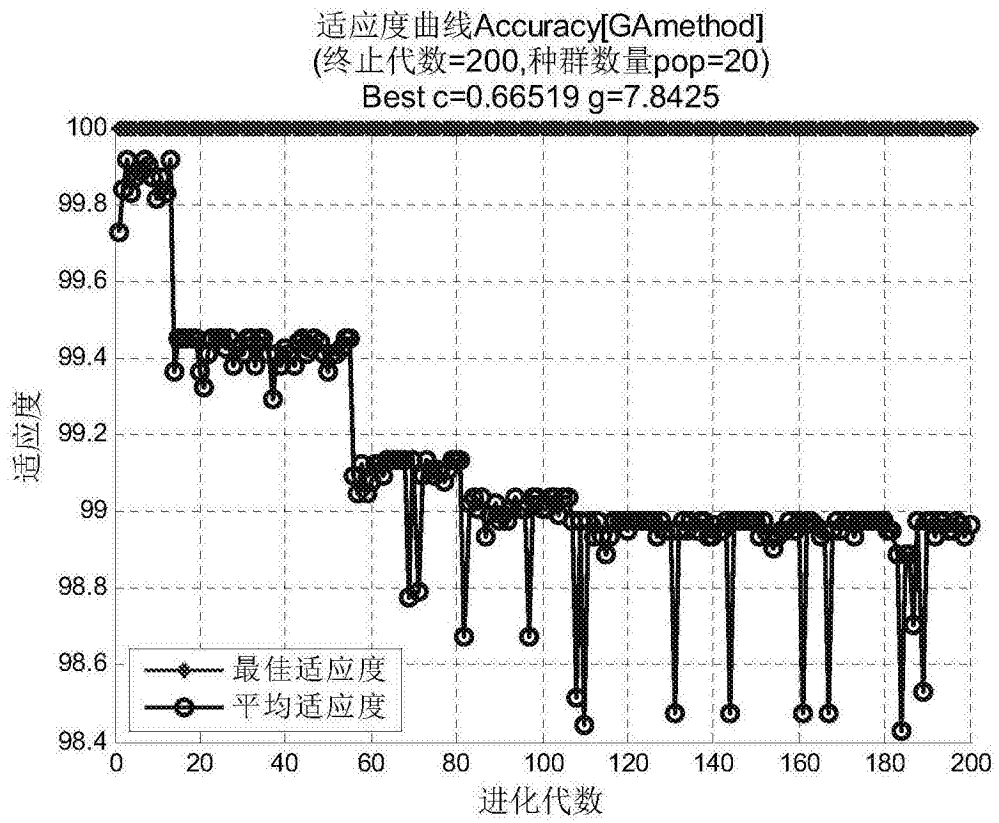


图5

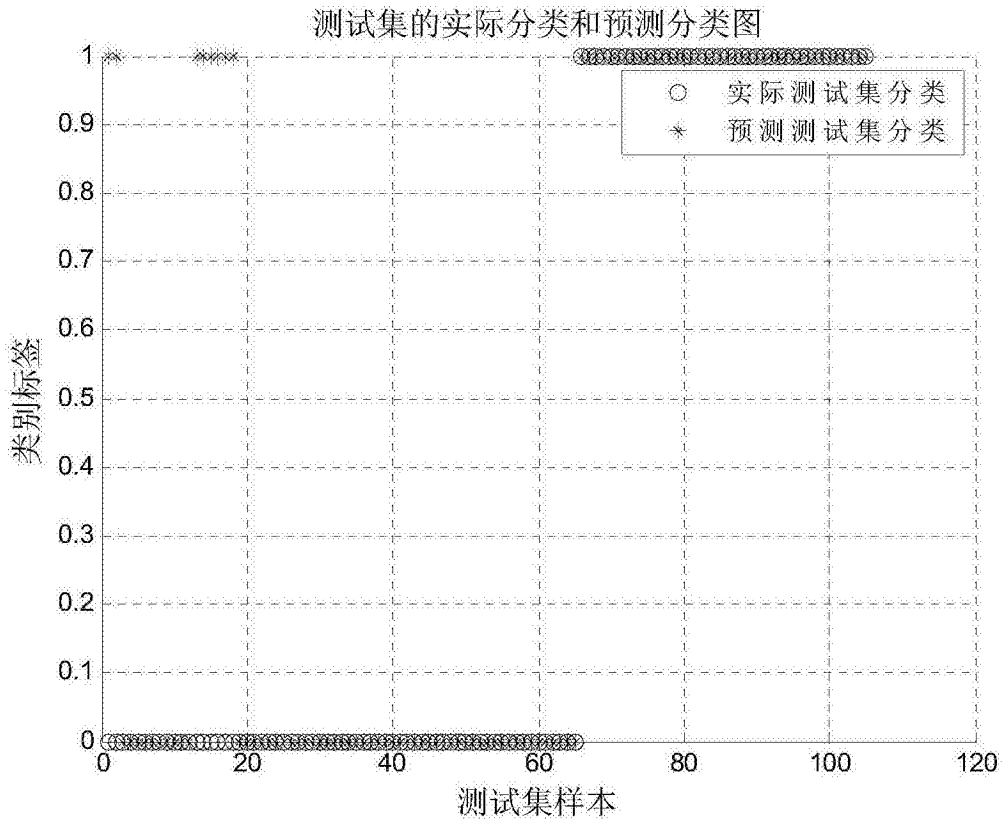


图6

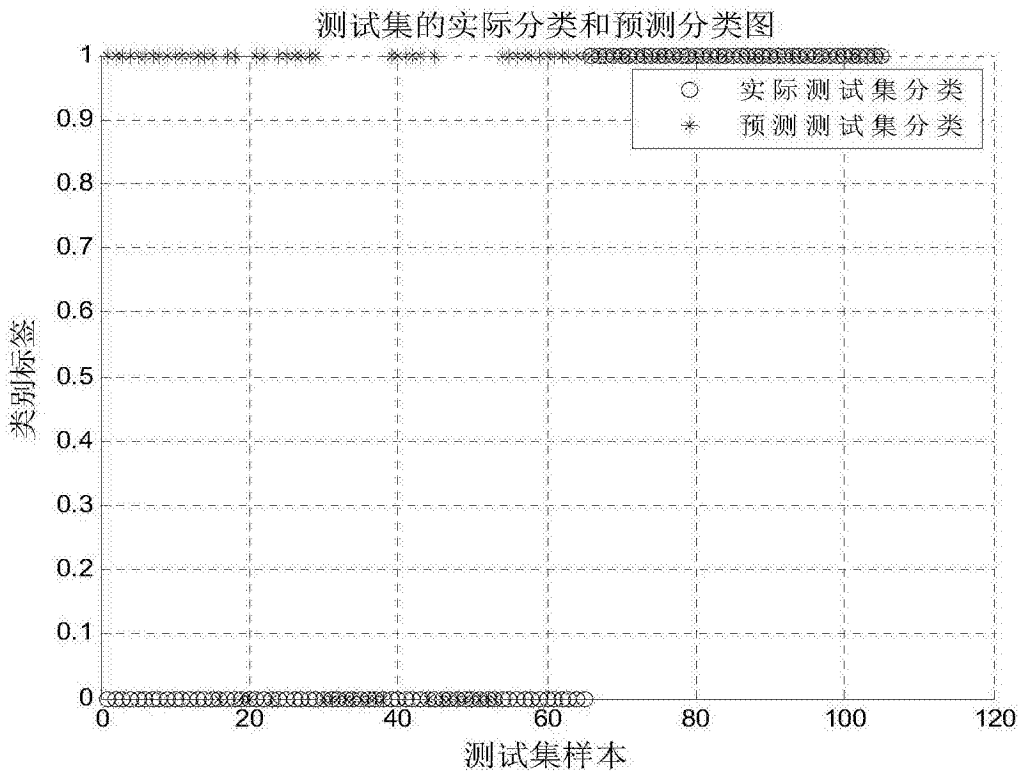


图7

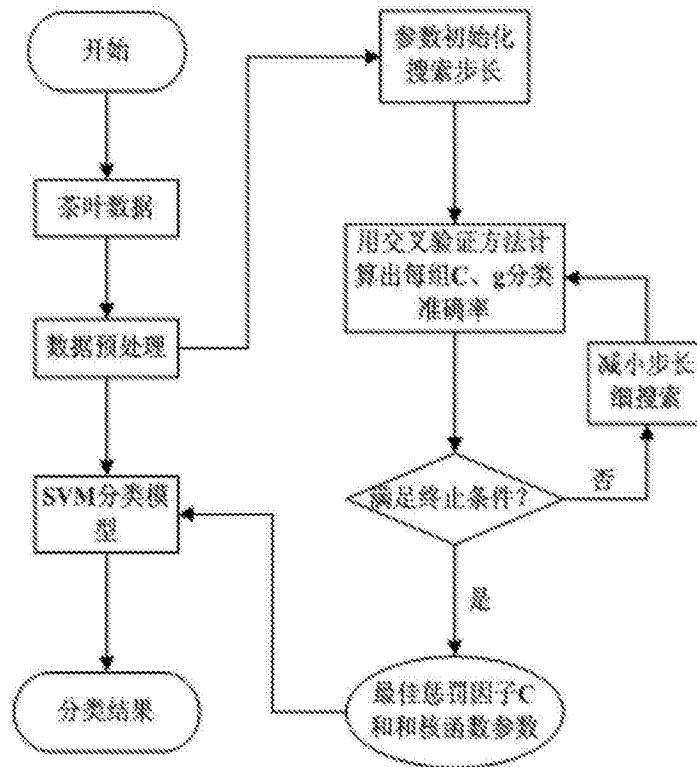


图8

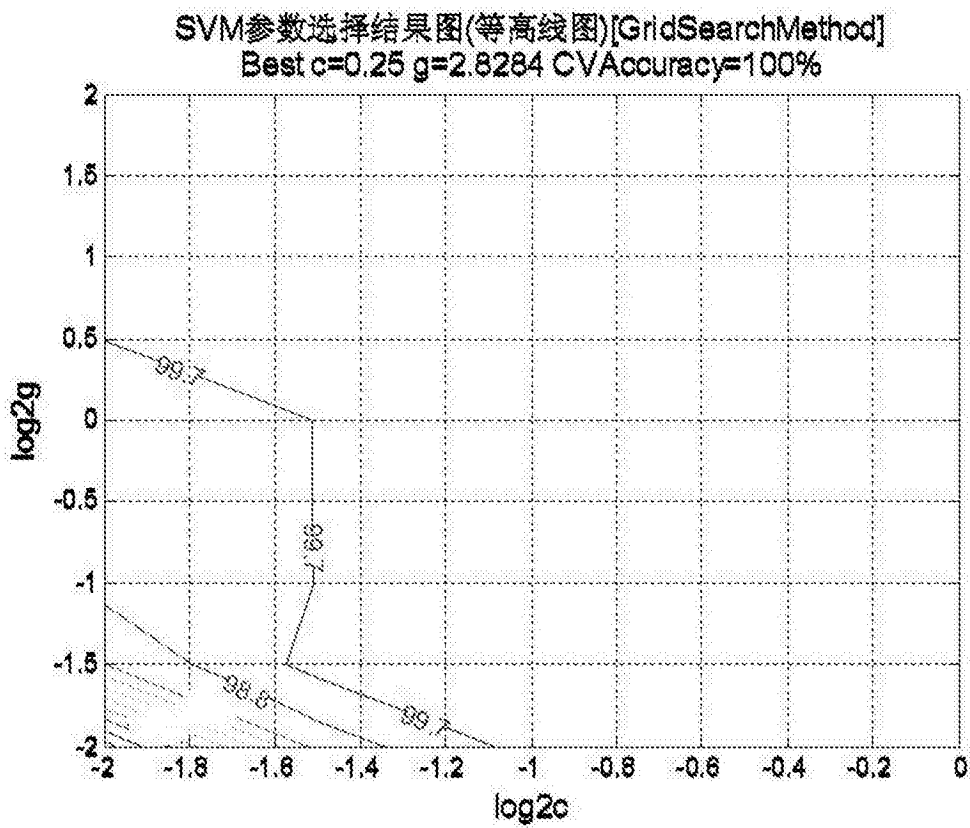


图9

SVM参数选择结果图(3D视图)[GridSearchMethod]
Best c=0.25 g=2.8284 CVAccuracy=100%

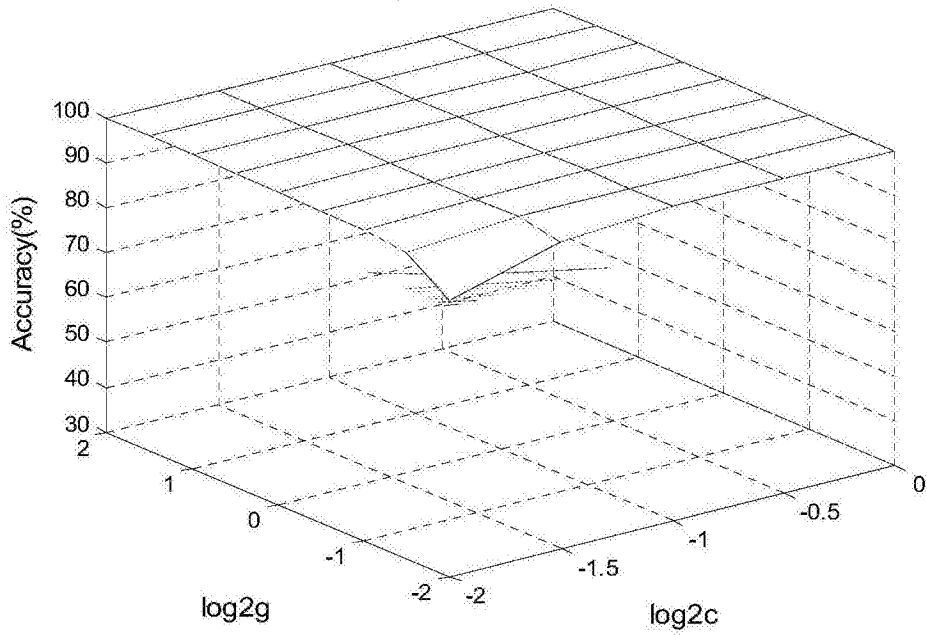


图10

测试集的实际分类和预测分类图

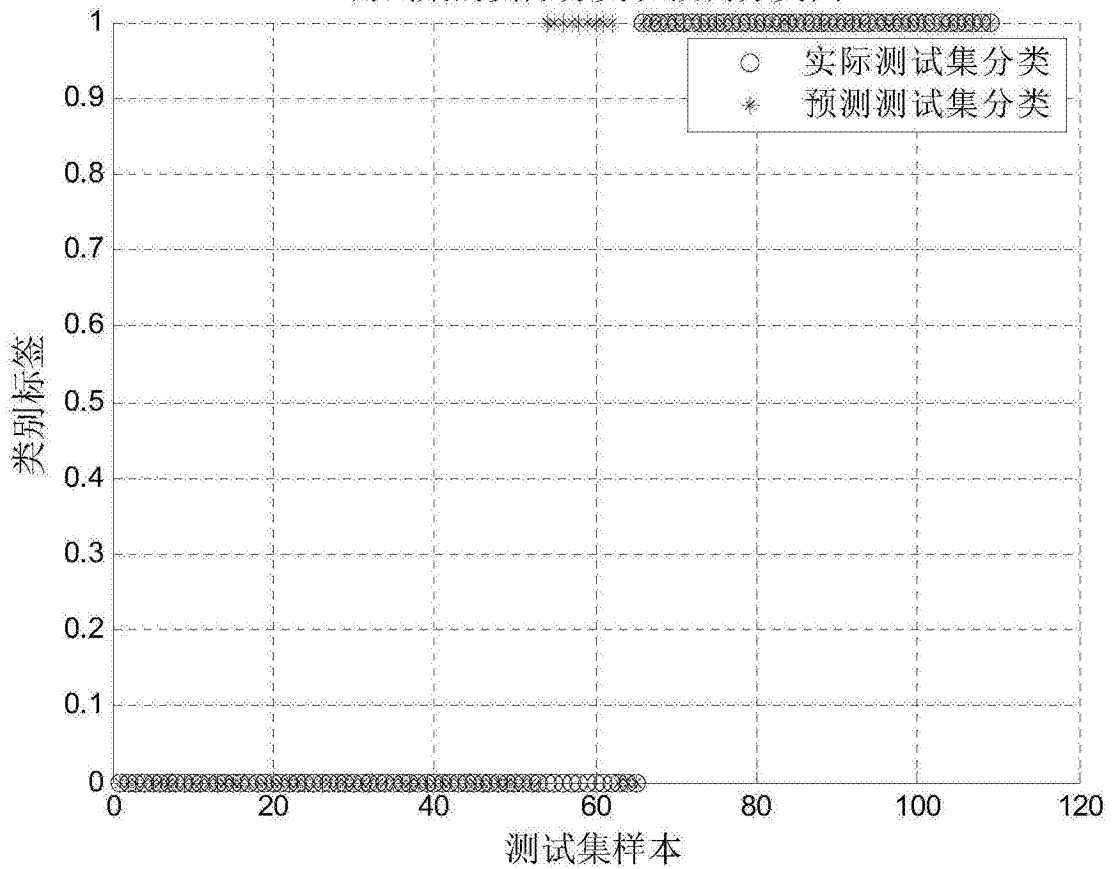


图11

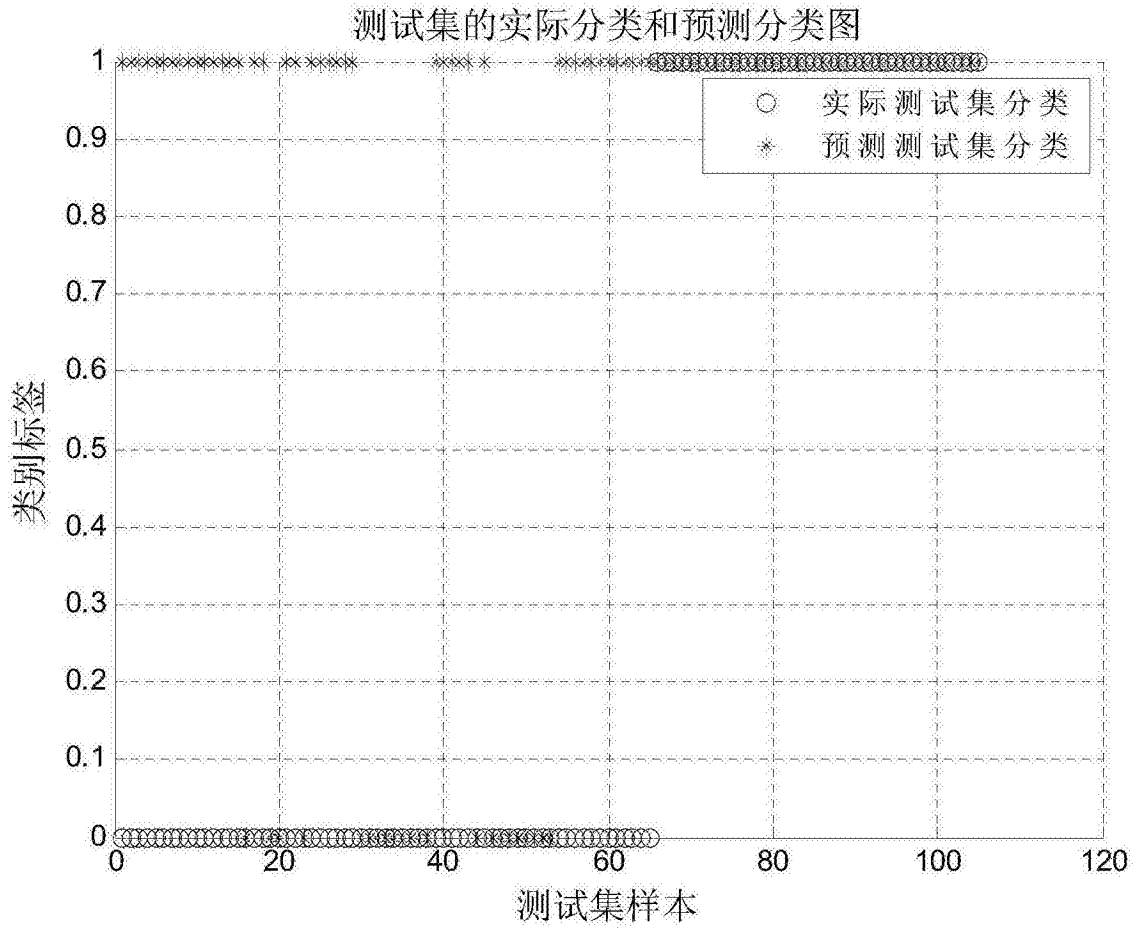


图12