



(12)发明专利申请

(10)申请公布号 CN 107480690 A

(43)申请公布日 2017.12.15

(21)申请号 201710536516.1

(22)申请日 2017.07.04

(71)申请人 中国科学院计算技术研究所

地址 100190 北京市海淀区中关村科学院
南路6号

(72)发明人 邢云冰 陈益强 忽丽莎

(74)专利代理机构 北京泛华伟业知识产权代理有限公司 11280

代理人 王勇

(51) Int.Cl.

G06K 9/62(2006.01)

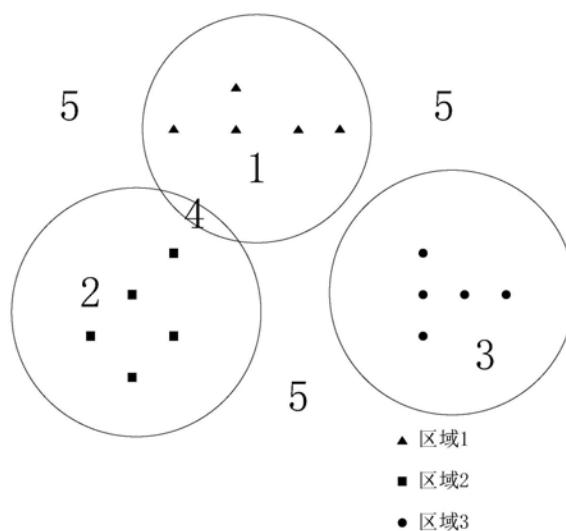
权利要求书2页 说明书6页 附图3页

(54)发明名称

一种基于支持向量机的包含未知类别的多分类方法

(57)摘要

本发明提供一种基于支持向量机的包含未知类别的多分类方法,包括:1)分别基于每个训练后的第i类别分类器,识别出待识别样本是否属于第i类别;所述第i类别分类器是用已知的属于第i类别的样本构建正样本集,用已知的属于其余N-1类别的所有样本构建负样本集,基于SVM模型进行训练而得到的用于识别所输入的样本是否属于第i类别的二分类分类器;2)当步骤1)中将待识别样本输入每个二分类分类器,所得结果均为否时,认定当前待识别样本属于未知类别;当步骤1)中,有且仅有一个二分类分类器的输出结果为是时,则认定当前待识别样本属于这个二分类分类器所对应的类别。本发明具有未知类别的检测能力,具有较高的查全率和查准率。



1. 一种基于支持向量机的包含未知类别的多分类方法,包括下列步骤:

1) 对于待识别样本,分别基于每个训练后的第*i*类别分类器,识别出该待识别样本是否属于第*i*类别;其中, $i=1,2,\dots,N$;所述第*i*类别分类器是用已知的属于第*i*类别的样本构建正样本集,用已知的属于其余*N*-1类别的所有样本构建负样本集,基于SVM模型进行训练而得到的二分类分类器,该二分类分类器用于识别所输入的样本是否属于第*i*类别;

2) 当步骤1) 中将待识别样本输入每个二分类分类器,所得结果均为否时,认定当前待识别样本属于未知类别;当步骤1) 中,有且仅有一个二分类分类器的输出结果为是时,则认定当前待识别样本属于这个二分类分类器所对应的类别。

2. 根据权利要求1所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述步骤2) 还包括:当步骤1) 中,有多个二分类分类器的输出结果为是时,直接认定当前待识别样本属于未知类别;或者在输出结果为是的这些二分类分类器所对应的类别的范围内,采用其它分类方法进一步地分辨当前待识别样本属于哪个类别。

3. 根据权利要求1所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述步骤1) 中,所述第*i*类别分类器按下述方法训练:

11) 取第*i*类别作为正类别,其余*N*-1种类别作为反类别;

12) 基于SVM模型进行训练而得到作为第*i*类别分类器的二分类分类器。

4. 根据权利要求3所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述步骤12) 中,在进行训练时,先将已知样本 \bar{x} 从原始空间映射到新的特征空间,得到对应的特征空间变换后的样本 $\phi(\bar{x})$,以使得已知样本的类别分界线更加规则。

5. 根据权利要求4所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述步骤12) 中,

SVM模型的目标函数被配置为:利用样本数据在特征空间中寻找一个超球面 $f(x) = \|\phi(\bar{x}) - \bar{O}\|^2 = R^2$,其中 \bar{O} 是超球面的球心, R 是超球面的半径,如果 $f(x) < R$ 则输出 $y = +1$ 表示该样本属于正类别,如果 $f(x) > R$ 则输出 $y = -1$,表示该样本属于反类别;

寻找最优的超球面,使得该超球面的表面积最小化并且已知样本数据离该超球面的最小距离最大化;然后根据所述的最优的超球面得到SVM模型的最优参数。

6. 根据权利要求5所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述步骤12) 中,使得已知样本数据离该超球面的最小距离最大化是:将正反类别样本数据离原超球面的最小距离分别形成两个同心超球面,使得同心超球面组成的超球壳的壳厚度最大化。

7. 根据权利要求5所述的基于支持向量机的包含未知类别的多分类方法,其特征在于,所述目标函数为:

$$G(\bar{O}, R, d, \bar{\xi}) = R^2 - 2C_1d + C_2 \sum_{i=0}^M \xi_i$$

约束条件为 $y_i \|\phi(\bar{x}_i) - \bar{O}\|^2 \leq y_i R^2 - d^2 + \xi_i$, $\xi_i \geq 0$,其中 C_1 和 C_2 是经验系数, d 相当于满足约束条件的样本数据离超球面的最小距离, ξ_i 为松弛变量;

所述的SVM模型的最优参数为： $\min_{\vec{O}, R, d, \vec{\xi}} G(\vec{O}, R, d, \vec{\xi})$ 。

8. 根据权利要求7所述的基于支持向量机的包含未知类别的多分类方法, 其特征在于, 所述步骤12) 中, 所述的SVM模型的最优参数求解方法如下:

121) 通过拉格朗日乘子法得到目标函数的拉格朗日函数;

122) 令拉格朗日函数中部分变量的偏导为零, 得到求解目标函数的对偶问题, 该对偶问题中的核函数选择高斯核或多项式核;

123) 通过二次规划算法求解所述对偶问题, 根据已知样本数据训练得到参数 \vec{O} 和 R 。

9. 根据权利要求7所述的基于支持向量机的包含未知类别的多分类方法, 其特征在于, 所述步骤12) 中, 所述特征空间变换能够使得SVM模型的类别线性可分或聚集可分。

一种基于支持向量机的包含未知类别的多分类方法

技术领域

[0001] 本发明涉及机器学习领域,具体地说,本发明涉及一种基于支持向量机的包含未知类别的多分类方法。

背景技术

[0002] 支持向量机(SVM)作为一种典型的机器学习算法,由于其完备的理论支持,在各种分类问题中得到了广泛应用。SVM假设样本数据 \bar{x}_i 在原始空间一般不是线性可分的,可以将样本数据从原始空间映射到新的特征空间(新特征空间一般是更高维的),样本数据 \bar{x}_i 在新特征空间的对应点为 $\phi(\bar{x}_i)$,二分类SVM的目标可以描述为利用样本数据在特征空间中寻找一个线性超平面 $f(x)=\bar{w}^T\phi(\bar{x})+b=0$,其中 \bar{w} 是线性超平面的法向量,b是偏移量,如果 $f(x_i)>0$ 则 $y_i=+1$ 表示该样本数据属于正类别,如果 $f(x_i)<0$ 则 $y_i=-1$ 表示该样本数据属于反类别。最优的线性超平面应该使得样本数据离该线性超平面的最小距离最大化,数学形式表示为 $\min_{\bar{w},b,\xi} \frac{1}{2}\|\bar{w}\|^2 + C\sum_{i=1}^M \xi_i$,满足约束条件 $y_i(\bar{w}^T\phi(\bar{x}_i)+b)\geq 1-\xi_i, \xi_i\geq 0$,其中C是经验系数, ξ_i 为松弛变量,对少数不满足硬性约束条件的样本数据放松约束(代价是目标函数被惩罚)。

[0003] 对于多分类SVM,任取两种类别的样本数据采用上述方式进行训练,分别得到分类模型(线性超平面),新样本数据最终属于哪个类别由所有分类模型投票产生。例如N分类SVM可以分解为 $N(N-1)/2$ 个二分类SVM,如果 $N(N-1)/2$ 个二分类SVM选取相同的核函数(特征空间映射函数)以及相关参数,相当于 $N(N-1)/2$ 个线性超平面将整个特征空间进行划分,如图1所示。

[0004] 目前几乎所有的机器学习算法,其判断结果均为已知类别,例如数字0-9的手写输入识别,无论输入何种数据,机器学习算法均会从这10个数字中选中其中一个作为识别结果,对于无效的随机输入数据,更理想的处理结果是输出一个未知类别来表示无效输出。在人机交互领域,该缺陷引起的问题并不明显,对于无效的识别结果,人类可以简单判别并忽略该输出。但在一些全自动领域,由于异常条件的存在,则必须自动判断输入数据是否正常或有效。

[0005] 对于SVM,目前可以采用如下三种方式解决该问题。

[0006] 第一种方式是在训练阶段获取未知类别的样本数据,对于数字识别即为采集各种随意输入的数据,例如波浪线(~~)、勾(√)、叉(×)等不属于数字0-9类别的数据,从而得到包含11种类别的样本数据,然后使用传统的SVM训练整个样本数据集得到分类模型。这种方式的缺点是明显的,未知类别的类别数本身就是无穷的,因此采集的样本数据无法覆盖所有分布。

[0007] 第二种方式是采用概率的形式,训练得到的是含有概率的分类模型,预测阶段输出的是每个新样本数据属于各种类别的概率而不是单一的类别结果。这种方式可以部分解

决该问题,尤其是预测阶段输出的前几个最大概率大小相当时,基本可以认定新样本数据是未知类别,因为属于已知类别时应该是对应该类别的概率值远远大于对应其他类别的概率值。如果某一已知类别和未知类别的某一子类别相对“接近”,该方式将会明显判断失误,例如数字1-9的识别,当实际输入“0”时,分类模型也将认为对应类别“6”的概率值远远大于对应其他类别的概率值。因此该方式是一个充分条件而非必要条件。

[0008] 第三种方式是联合使用SVM的变种算法——支持向量数据描述 (SVDD) 算法。在训练阶段,首先将所有已知类别的样本数据作为正常类别,其次利用这一正常类别的样本数据训练得到SVDD模型,然后使用传统的SVM训练整个样本数据集得到SVM模型。在预测阶段,首先根据SVDD 模型判断新样本数据是否正常,如果正常,则使用SVM模型判断新样本数据具体属于哪一个类别。这种方式的主要缺点存在于SVDD阶段,由于只有一种类别的样本数据,训练得到的SVDD模型一般较为粗糙,同时缺少其他类别样本数据的限制和矫正,SVDD模型也容易产生过拟合,因此对于新样本数据是否正常的判断精度本身就较差,亦即对新样本数据是否属于未知类别的判断精度较差。

发明内容

[0009] 因此,本发明的任务是提供一种更加适于含有未知类别的应用场景的基于支持向量机的多分类解决方案。

[0010] 根据本发明的一个方面,提供了一种基于支持向量机的包含未知类别的多分类方法,包括下列步骤:

[0011] 1) 对于待识别样本,分别基于每个训练后的第*i*类别分类器,识别出该待识别样本是否属于第*i*类别;其中, $i=1,2,\dots,N$;所述第*i*类别分类器是用已知的属于第*i*类别的样本构建正样本集,用已知的属于其余*N-1* 类别的所有样本构建负样本集,基于SVM模型进行训练而得到的二分类分类器,该二分类分类器用于识别所输入的样本是否属于第*i*类别;

[0012] 2) 当步骤1) 中将待识别样本输入每个二分类分类器,所得结果均为否时,认定当前待识别样本属于未知类别;当步骤1) 中,有且仅有一个二分类分类器的输出结果为是时,则认定当前待识别样本属于这个二分类分类器所对应的类别。

[0013] 其中,所述步骤2) 还包括:当步骤1) 中,有多个二分类分类器的输出结果为是时,直接认定当前待识别样本属于未知类别;或者在输出结果为是的这些二分类分类器所对应的类别的范围内,采用其它分类方法进一步地分辨当前待识别样本属于哪个类别。

[0014] 其中,所述步骤1) 中,所述第*i*类别分类器按下述方法训练:

[0015] 11) 取第*i*类别作为正类别,其余*N-1*种类别作为反类别;

[0016] 12) 基于SVM模型进行训练而得到作为第*i*类别分类器的二分类分类器。

[0017] 其中,所述步骤12) 中,在进行训练时,先将已知样本 \bar{x} 从原始空间映射到新的特征空间,得到对应的特征空间变换后的样本 $\phi(\bar{x})$,以使得SVM 模型的类别分界线更加规则。

[0018] 其中,所述步骤12) 中,

[0019] SVM模型的目标函数被配置为:利用样本数据在特征空间中寻找一个超球面 $f(x) = \|\phi(\bar{x}) - \bar{O}\|^2 = R^2$,其中 \bar{O} 是超球面的球心, R 是超球面的半径,如果 $f(x) < R$ 则输出 $y = +1$ 表示该样本属于正类别,如果 $f(x) > R$ 则输出 $y = -1$,表示该样本属于反类别;

[0020] 寻找最优的超球面,使得该超球面的表面积最小化并且已知样本数据离该超球面的最小距离最大化;然后根据所述的最优的超球面得到SVM模型的最优参数。

[0021] 其中,所述步骤12)中,使得已知样本数据离该超球面的最小距离最大化是:将正反类别样本数据离原超球面的最小距离分别形成两个同心超球面,使得同心超球面组成的超球壳的壳厚度最大化。

[0022] 其中,所述目标函数为:

$$[0023] \quad G(\vec{O}, R, d, \vec{\xi}) = R^2 - 2C_1 d + C_2 \sum_{i=0}^M \xi_i$$

[0024] 约束条件为 $y_i \|\phi(\vec{x}_i) - \vec{O}\|^2 \leq y_i R^2 - d^2 + \xi_i$, $\xi_i \geq 0$, 其中 C_1 和 C_2 是经验系数, d 相当于满足约束条件的样本数据离超球面的最小距离, ξ_i 为松弛变量;

[0025] 所述的SVM模型的最优参数为: $\min_{\vec{O}, R, d, \vec{\xi}} G(\vec{O}, R, d, \vec{\xi})$ 。

[0026] 其中,所述步骤12)中,所述的SVM模型的最优参数求解方法如下:

[0027] 121) 通过拉格朗日乘子法得到目标函数的拉格朗日函数;

[0028] 122) 令拉格朗日函数中部分变量的偏导为零,得到求解目标函数的对偶问题,该对偶问题中的核函数优选高斯核;

[0029] 123) 通过二次规划算法求解所述对偶问题,根据已知样本数据训练得到参数 \vec{O} 和 R 。需要说明,在步骤122)中,虽然在大部分情况高斯核是最优的,但高斯核并不是唯一的选择,在一些实施例,多项式核(即用多项式核作为核函数)对部分场景也是有效的。

[0030] 其中,所述步骤12)中,所述特征空间变换能够使得SVM模型的类别线性可分或聚集可分。

[0031] 与现有技术相比,本发明具有下列技术效果:

[0032] (1) 同类别的样本数据是可聚类的,而非发散的,符合数据的实际分布,因此本发明更贴近实际数据。判断未知类别的精度较高。

[0033] (2) 分类模型具有未知类别的检测能力,不仅具有和传统SVM相当的查全率而且还具有较高的查准率。

[0034] (3) 分类模型简单,类似于传统的SVM,最终分类模型仅和少量位于超球壳界面上或者错误侧的样本数据(即支持向量)有关。

[0035] (4) 预测阶段算法简单,实现复杂度低,特别适合于模型相对稳定的一次训练,多次预测的场合。

附图说明

[0036] 以下,结合附图来详细说明本发明的实施例,其中:

[0037] 图1是现有技术中的传统SVM的三分类方案示意图;

[0038] 图2是本发明一个实施例中的一个二分类方案的原理图;

[0039] 图3是本发明一个实施例中的三分类方案的示意图。

具体实施方式

[0040] 下面结合附图和具体实施方式对本发明加以说明。

[0041] 根据本发明的一个实施例,提供了一种包含未知类别的三分类方法。该三分类方法可以分解为三个二分类方法,每个二分类方法包括训练和预测两个阶段。

[0042] (一) 训练阶段:

[0043] 步骤1)、选取其中一种类别作为正类别,其余两种类别作为反类别。

[0044] 步骤2)、将样本数据 \bar{x}_i 从原始空间映射到新的特征空间,样本数据 \bar{x}_i 在新特征空间的对应点为 $\phi(\bar{x}_i)$,如图2和图3中的▲、■和●所示。特征空间映射是SVM分类算法中普遍采取的一种数据预处理方式,其目的是使得样本数据在新特征空间(映射后特征空间维度更高)中更容易分开,比如线性可分或聚集可分。也就是说,本步骤中,特征空间映射能够让不同样本类别之间的分界线更规则些。本实施例中,如果能够用一个线性函数可以将两类样本完全分开,就称这些样本线性可分。如果能够用一个对应于超球面的函数将两类样本完全分开,就称这些样本是聚集可分的。在具体实现上,特征空间映射可以是从小- y 特征空间到 r - θ 特征空间的映射,其特征空间变换公式为:

$$[0045] \quad r = \sqrt{x^2 + y^2}$$

$$[0046] \quad \theta = \arctan(y, x)$$

[0047] 特征空间映射也可以是从 x - y 特征空间到 u - v 特征空间的映射,其特征空间变换公式为:

$$[0048] \quad u = x^{-1}$$

$$[0049] \quad v = y^{-1}$$

[0050] 上述两种特征空间映射方式仅仅是示例性的,在其他实施例中,也可以采用其他的特征空间映射方式,只要能够使已知样本的类别分界线更加规则即可。

[0051] 步骤3)、二分类方法的目标函数可以描述为:利用样本数据在特征空间中寻找一个超球面 $f(x) = \|\phi(\bar{x}) - \bar{O}\|^2 = R^2$,其中 \bar{O} 是超球面的球心, R 是超球面的半径,如果 $f(x) < R$ 则 $y_i = +1$ 表示该样本属于正类别,如果 $f(x) > R$ 则 $y_i = -1$ 表示该样本属于反类别。最优的超球面应该使得超球面的表面积最小化并且样本数据离该超球面的最小距离最大化,即该超球面的表面积尽量小,正反类别样本数据离原超球面的最小距离分别形成两个同心超球面,同心超球面组成的超球壳的壳厚度尽量大,如图2所示,数学形式表达为

$$[0052] \quad \min_{\bar{O}, R, d, \xi} R^2 - 2C_1 d + C_2 \sum_{i=0}^M \xi_i \quad (1)$$

[0053] 满足约束条件 $y_i \|\phi(\bar{x}_i) - \bar{O}\|^2 \leq y_i R^2 - d^2 + \xi_i$, $\xi_i \geq 0$,其中 C_1 和 C_2 是经验系数, d 相当于满足约束条件的样本数据离超球面的最小距离, ξ_i 为松弛变量,对少数不满足硬性约束条件的样本数据放松约束(代价是目标函数被惩罚)。

[0054] 步骤4)、通过拉格朗日乘子法得到式1的拉格朗日函数

[0055]

$$L(\bar{O}, R, d, \bar{\alpha}, \bar{\xi}, \bar{\mu}) = R^2 - 2C_1 d + C_2 \sum_{i=1}^M \xi_i + \sum_{i=1}^M \alpha_i \left(y_i \|\phi(\bar{x}_i) - \bar{O}\|^2 - y_i R^2 + d^2 - \xi_i \right) - \sum_{i=1}^M \mu_i \xi_i \quad (2)$$

[0056] 其中 $\alpha_i \geq 0$, $\mu_i \geq 0$ 是拉格朗日乘子。

[0057] 步骤5)、令 $L(\bar{O}, R, d, \bar{\alpha}, \bar{\xi}, \bar{\mu})$ 对 \bar{O} 、 R 、 d 和 $\bar{\xi}$ 的偏导为零, 可得到式 (1) 的对偶问题

$$[0058] \quad \max_{\bar{\alpha}} \sum_{i=1}^M \alpha_i y_i K(\bar{x}_i, \bar{x}_i) - \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j K(\bar{x}_i, \bar{x}_j) \quad (3)$$

[0059] 其中 $K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i)^T \phi(\bar{x}_j)$ 表示核函数, 满足约束条件 $\sum_{i=1}^M \alpha_i y_i = 1$, $0 \leq \alpha_i \leq C_2$ 。优选

地, 核函数可以选择高斯核 $K(\bar{x}_i, \bar{x}_j) = \exp\left(-\frac{\|\bar{x}_i - \bar{x}_j\|^2}{2\sigma^2}\right)$ 。

[0060] 步骤6)、式 (3) 是一个典型的二次规划问题, 使用通用的二次规划算法来求解。优选的, 可以使用序列最小化优化 (SMO) 算法对式 (3) 求解, 通过样本数据训练得到参数 \bar{O} 和 R , 即得到一个分类模型。

[0061] (二) 预测阶段

[0062] 对于每个新的样本数据, 根据分类模型预测 $f(x) = \|\phi(\bar{x}) - \bar{O}\|^2$, 如果 $f(x) < R$ 则表示该样本属于正类别, 如果 $f(x) > R$ 则表示该样本属于反类别。

[0063] 对于本实施例的三分类方法, 选取其中每一种类别作为正类别, 其余所有类别作为反类别, 采用上述二分类方式进行训练, 分别得到分类模型, 新样本数据根据每个分类模型分别预测, 最终可以判断是否属于已知类别以及属于哪个已知类别。

[0064] 如果三个二分类问题选取相同的核函数以及相关参数, 相当于三个超球面将整个特征空间划分为四个部分, 如果新样本数据位于所有三个超球面的外部, 则认为该样本属于未知类别, 如果新样本数据位于一个以上超球面的内部 (多个超球面相交), 可以采用其他方式判断该样本的类别, 例如传统的 SVM, 也可以直接认为该样本属于未知类别, 如图3所示。

[0065] 本发明的效果可以通过以下实验说明。

[0066] 本实验比较传统 SVM 和本发明的分类效果。传统 SVM 的参数配置为: 线性核函数 (无特征空间映射), 经验参数 $C=1$; 本发明的参数配置为: 线性核函数 (无特征空间映射), 经验参数 $C_1=10, C_2=1$ 。

[0067] 本实验所用数据来源于日常行为数据集中的红外数据, 详见“Jiang X, Chen Y, Liu J, et al. AIR: recognizing activity through IR-based distance sensing on feet[C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM, 2016: 97-100”。该数据集包含6类共3372个样本数据, 分别为走 (walk)、下楼 (downstair)、上楼 (upstair)、跑 (run)、站 (stay) 和原地走 (walk_in_place), 每个样本数据为38维。

[0068] 本实验利用上述数据集中的前5类数据作为训练数据, 所有6类数据作为测试数据。实验环境为 WIN7, Matlab 2016b, 优化算法 (SMO) 使用 CVX 工具。表1示出了传统 SVM 的分类结果。

[0069] 表1

[0070]

| | | 预测行为 | | | | | | 查全率 |
|------|----|--------|--------|--------|--------|--------|----|----------------|
| | | 走 | 下楼 | 上楼 | 跑 | 站 | 未知 | |
| 真实行为 | 走 | 779 | 1 | 2 | 3 | 0 | 0 | 99.24% |
| | 下楼 | 31 | 394 | 0 | 0 | 2 | 0 | 92.27% |
| | 上楼 | 15 | 3 | 438 | 0 | 0 | 0 | 96.05% |
| | 跑 | 0 | 0 | 0 | 529 | 0 | 0 | 100% |
| | 站 | 0 | 0 | 0 | 0 | 622 | 0 | 100% |
| | 未知 | 0 | 314 | 239 | 0 | 0 | 0 | 0% |
| 查准率 | | 94.42% | 55.34% | 64.51% | 99.44% | 99.68% | 0% | 81.91% (精度) |

[0071] 表2示出了本发明一个实施例的SVM分类结果。

[0072] 表2

[0073]

| | | 预测行为 | | | | | | 查全率 |
|------|----|--------|--------|--------|--------|------|--------|----------------|
| | | 走 | 下楼 | 上楼 | 跑 | 站 | 未知 | |
| 真实行为 | 走 | 780 | 2 | 0 | 1 | 0 | 2 | 99.36% |
| | 下楼 | 47 | 366 | 1 | 0 | 0 | 13 | 85.71% |
| | 上楼 | 27 | 8 | 392 | 0 | 0 | 29 | 85.96% |
| | 跑 | 5 | 4 | 5 | 512 | 0 | 3 | 96.79% |
| | 站 | 0 | 0 | 0 | 0 | 616 | 6 | 99.04% |
| | 未知 | 2 | 10 | 10 | 0 | 0 | 531 | 96.02% |
| 查准率 | | 90.59% | 93.85% | 96.08% | 99.81% | 100% | 90.92% | 94.81% (精度) |

[0074] 实验结果表明,本发明使分类模型具有未知类别的检测能力,不仅具有和传统支持向量机相当的查全率而且具有较高的查准率,从而使得分类的整体精度也得到提高。

[0075] 最后所应说明的是,以上实施例仅用以说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,对本发明的技术方案进行修改或者等同替换,都不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

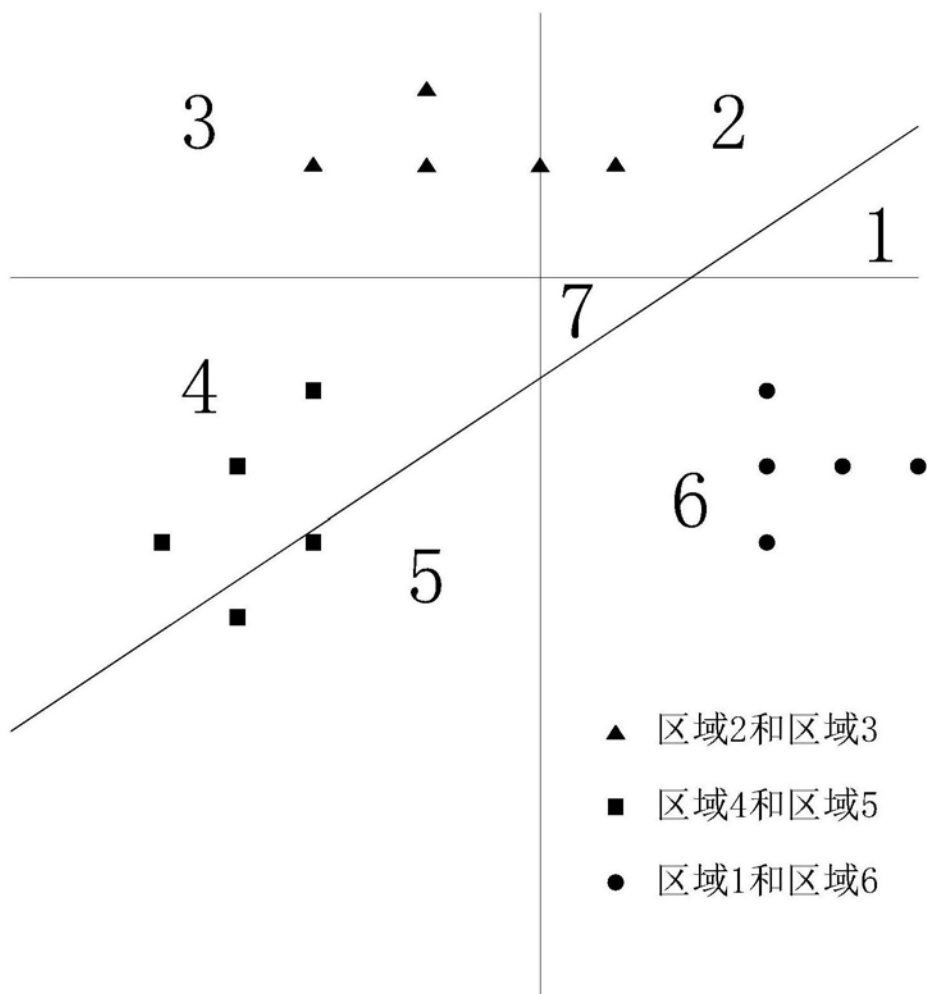


图1

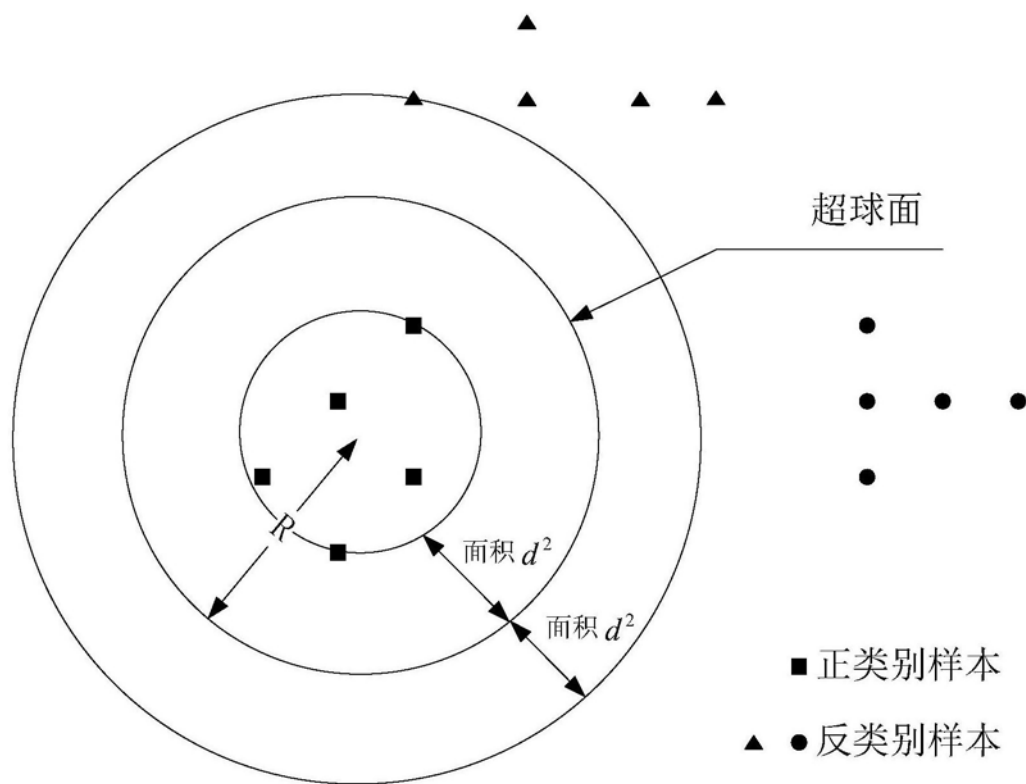


图2

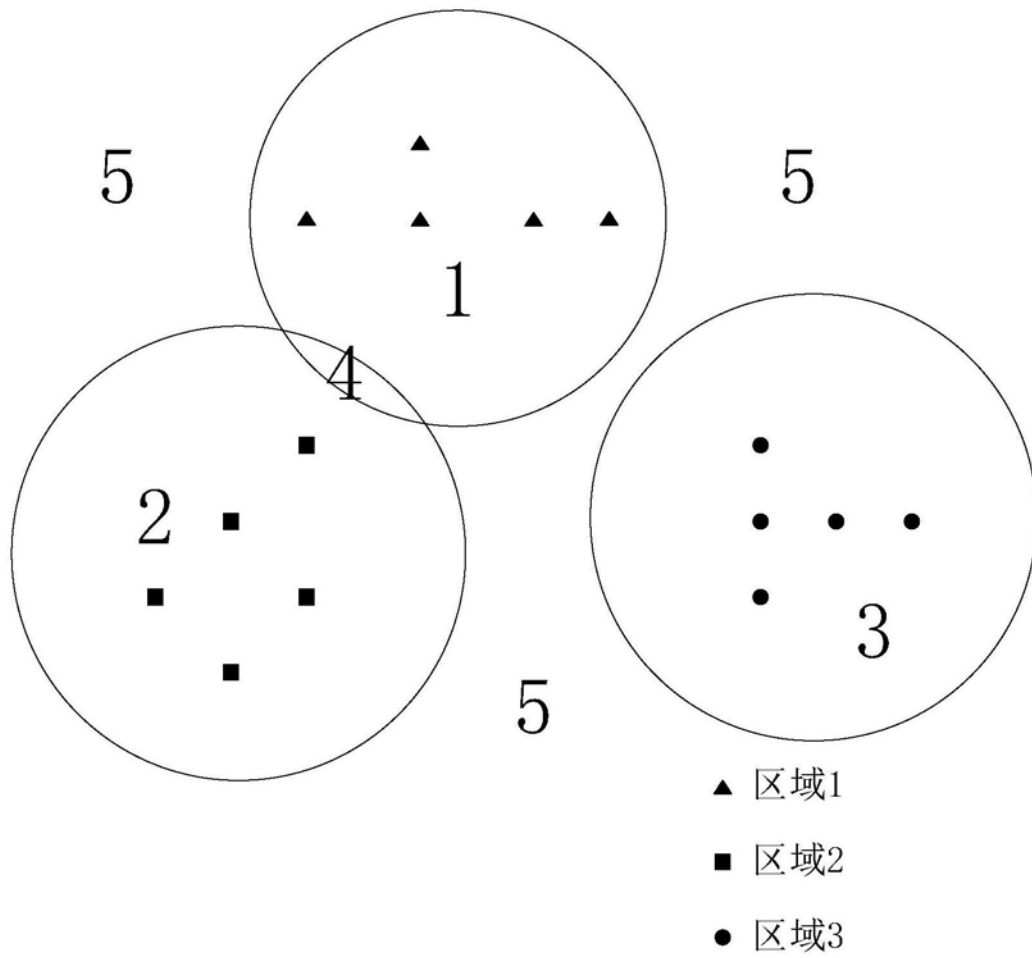


图3