(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108564117 A
(43)申请公布日 2018.09.21

(21)申请号 201810290654.0

(22)申请日 2018.03.30

(71)申请人 华南理工大学
地址 510640 广东省广州市天河区五山路381号

(72)发明人 彭新一 余珍

(74)专利代理机构 广州市华学知识产权代理有限公司 44245
代理人 陈宏升

(51)Int.Cl.
G06K 9/62(2006.01)
G06Q 50/20(2012.01)

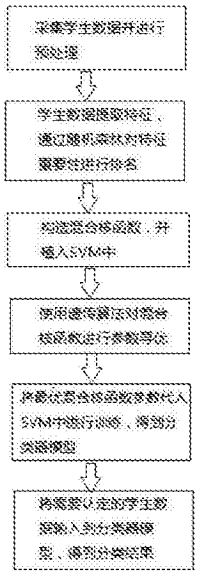
权利要求书3页 说明书8页 附图1页

(54)发明名称

一种基于SVM的贫困生辅助认定方法

(57)摘要

本发明公开了一种基于SVM的贫困生辅助认定方法，包括以下步骤：采集学生数据，并对学生数据进行预处理；对预处理学生数据提取特征，通过随机森林对特征重要性进行排名；在满足Mercer条件下，构造混合核函数，并植入SVM中；使用遗传算法对混合核函数参数进行寻优，得到最优混合核函数参数；将最优混合核函数参数代入SVM，并对学生数据进行训练，得到分类器模型；将需要认定的学生数据输入分类器模型，输出分类结果；本发明采用遗传算法对基于SVM混合核函数参数进行寻优，用适应度作为评价依据，通过随机重组重要基因，让群体中的个体不断进化，获取最优解，减少全局搜索时间，提高了分类器的推广泛化能力，并降低成本。



CN 108564117 A

1.一种基于SVM的贫困生辅助认定方法，其特征在于，包括以下步骤：

S1、采集学生数据，并对学生数据进行预处理；

S2、对预处理学生数据提取特征，通过随机森林对特征重要性进行排名；

S3、在满足Mercer条件下，构造混合核函数，并植入SVM中；

S4、使用遗传算法对混合核函数参数进行寻优，得到最优混合核函数参数；

S5、将最优混合核函数参数代入SVM中进行训练，训练之后得到分类器模型；

S6、将需要认定的学生数据输入分类器模型，输出分类结果。

2.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，步骤S1中，所述学生数据包含学生一卡通流水记录、学生基本信息、学生成绩和贫困生名单；所述学生基本信息包含学生ID、学生性别、学生名字。

3.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，步骤S1中，所述预处理包含去重、缺失值处理和格式化；

所述去重为：将学生数据按学生ID进行排序，通过比较邻近记录是否相似来检测记录是否重复，重复则删除重复记录；

所述缺失值处理为：学生数据中某个记录的某个字段为空，则使用平均值进行填充；

所述格式化为：将消费时间格式化为yyyy-MM-dd；消费金额统一单位为分，超限则四舍五入。

4.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，所述步骤S2具体过程为：

U1、从学生一卡通流水记录构造特征；从时间维度、地点维度和交易维度统计均值和方差；

U2、将学生一卡通流水记录数据特征、学生基本信息数据特征、学生成绩数据特征和贫困生名单数据特征，进行归一化；

U3、使用随机森林对特征重要性进行排名，根据排名，选择前30个特征。

5.根据权利要求4所述的一种基于SVM的贫困生辅助认定方法，其特征在于，所述使用随机森林对特征重要性进行排名具体为：

Y1、设定N个样本，每个样本有M个特征；

Y2、从N个样本中有放回的随机抽取，抽取N次，作为训练一棵决策树的样本；

Y3、每个节点随机抽取m个特征，m＜M，从中选取信息增益最大的特征作为决策树的分裂节点，在决策树成长的过程中，m值保持不变；

Y4、重复步骤Y2、Y3，建立大量的决策树，构成随机森林；

Y5、计算每个特征在随机森林中每棵树上的评分均值，作为特征重要性依据。

6.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，所述步骤S3，具体过程为：

基于对局部核函数和全局核函数，构造混合核函数，并植入SVM中：

$$K(x_i,x_j)=(1-\rho)\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)+\rho\left[(x_i\cdot x_j)+c\right]^d,0<\rho<1,$$

其中，ρ为混合核函数权系数，$\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)$为高斯核，属于局部核函数；σ为高斯核的带宽，σ＞0，$[(x_i \cdot x_j)+c]^d$为多项式核，属于全局核函数，c为自由参数，c≥0；d为多项式次数，d≥1，x_i为第i个样本的特征值向量，x_j为第j个样本的特征值向量。

7.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，步骤S4中，所述寻优过程具体如下：

V1、设置参数：初始种群数量为60，选择代购为0.8，交叉概率为0.6，变异概率为0.06；

V2、使用遗传算法确认混合核函数最优混合核函数参数，确认惩罚因子和确认混合核函数权系数；

V3、混合核函数参数、混合核函数权系数和惩罚因子采用二进制编码，并把其二进制编码组合得到个体染色体基因串，构造出多个染色体组合一个初始种群；

V4、根据初始种群计算适应度值：

$$P=\frac{TP}{TP+FP},$$

$$R=\frac{TP}{TP+FN},$$

其中，P为查准率，R为查全率，TP为真正例数目，FP为假正例数目，FN为假反例数目；

设遗传算法中的适应度值为$f(X_i)$，即10折交叉验证的macroF1值，则有：

$$\mathrm{macro}P=\frac{1}{10}\sum_{i=1}^{10}P_i,$$

$$\mathrm{macro}R=\frac{1}{10}\sum_{i=1}^{10}R_i,$$

$$\mathrm{macroF1}=\frac{2*macroP*macroR}{macroP+macroR},$$

其中，P_i为第i次训练查准率；macroP为宏查准率，是10次训练查准率平均值；R_i为第i次训练查全率；macroR为宏查全率，是10次训练查全率平均值；macroF1为宏F1，是基于宏查准率和宏查全率的调和平均值，即为适应度值；

V5、根据适应度值计算染色体入选种群概率：

$$p(X_i)=\frac{f(X_i)}{\sum f(X_j)},$$

其中，$p(X_i)$为第i个染色体入选种群概率；X_i为第i个染色体；$f(X_j)$为第j个染色体的适应度值；

V6、根据入选种群概率的高低，选择代沟为0.8，即保留概率较高的80％染色体，将保留的染色体进行交叉运算和变异运算：

所述交叉运算为随机选取两条染色体，随机选择一个交配点做单点杂交，将产生的新的两条染色体代替原来的染色体，放回初始种群；交叉运算概率为0.6；

所述变异运算为杂交后的个体进行变异运算，随机选取一条染色体，该染色体某个二进制位有6％的概率变异，即由0变1或由1变0；

V7、通过不断进化，获取最优混合核函数系数、最优确认惩罚因子和最优确认混合核函数权系数，从而确定混合核函数。

8.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，所述步骤S5，具体过程为：

根据步骤S4获得的最优混合核函数系数，使用SMO算法通过训练学生数据得到最优的$\hat{a}$、$\hat{b}$，其中，$\hat{a}$为拉格朗日乘子的最优解，$\hat{b}$为分类超平面的位移最优解；即：SMO每次选取两个拉格朗日乘子，固定其余参数；求解：

$$\min L(a)=\frac{1}{2}\sum_{i,j=1}^{l}a_i a_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{l}a_i ,$$

$$s.t\ \sum_{i=1}^{l}a_i y_i=0, 0\le a_i\le C, i=1,2,\ldots,l ,$$

其中，a_i、a_j为拉格朗日乘子；y_i为第i个学生标识，y_j为第j个学生标识；

获得更新后的a_i、a_j；

求解非线性支持向量机和其对偶问题，重复选取和求解，得到$\hat{a}$、$\hat{b}$；

其中非线性支持向量机为：

$$\begin{cases} \min_{\omega,b}\ \frac{1}{2}\omega^2+C\sum_{i=1}^{l}\xi_i \\ s.t\ y_i\left(\omega\cdot\Phi\left(x_i\right)+b\right)\ge 1-\xi_i, i=1,2,\ldots,l \end{cases},$$

对偶问题：

$$\begin{cases} \min L\left(\alpha\right)=\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(x_i,x_j\right)-\sum_{i=1}^{l}\alpha_i \\ s.t\ \sum_{i=1}^{l}\alpha_i y_i=0, 0\le\alpha_i\le C, i=1,2,\ldots,l \end{cases},$$

通过$\hat{a}$、$\hat{b}$，得到分类器模型：

$$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{l}\hat{a}_i y_i K(x_i,x)+\hat{b}\right),$$

其中，x为需要认定的学生数据特征值。

9.根据权利要求1所述的一种基于SVM的贫困生辅助认定方法，其特征在于，所述步骤S6，具体如下：

将需要认定的学生数据输入到分类器模型中，通过分类器模型，得出f(x)，若为正则表示这个学生大概率为贫困生，若为负则表示这个学生大概率不是贫困生，再通过实际考核，认定新的贫困生，添加到贫困生名单中，得到新的贫困生名单。

一种基于SVM的贫困生辅助认定方法

技术领域

[0001] 本发明涉及SVM核函数研究领域，特别涉及一种基于SVM的贫困生辅助认定方法。

背景技术

[0002] 随着高等教育的发展，越来越多贫困生进入大学，资助贫困生也成为高校重要的学生工作。而贫困生资格认定是高校资助工作的前提和基础。

[0003] 目前主流的认定方式是通过人工甄别申请材料，认定过程中存在认定程序僵化、责任主体缺乏伦理监督等问题，难以保证公平客观公正。在信息爆炸时代，兴起的机器学习方法尚不能提出很好的解决方案，在分类器的训练上、分类器的拟合上都存着各种各样的问题。基于统计学习理论提出的支持向量机SVM 遵循结构风险最小化原则，有效地避免了维数灾难，但其算法训练时间复杂度较高，泛化能力不够理想，在贫困生辅助认定的应用中始终乏力。

发明内容

[0004] 本发明的主要目的在于克服现有技术的缺点与不足，提供一种基于SVM的贫困生辅助认定方法。

[0005] 本发明的目的通过以下的技术方案实现：

[0006] 一种基于SVM的贫困生辅助认定方法，包括以下步骤：

[0007] S1、采集学生数据，并对学生数据进行预处理；

[0008] S2、对预处理学生数据提取特征，通过随机森林对特征重要性进行排名；

[0009] S3、在满足Mercer条件下，构造混合核函数，并植入支持向量机SVM中；

[0010] S4、使用遗传算法对混合核函数参数进行寻优，得到最优混合核函数参数；

[0011] S5、将最优混合核函数参数代入SVM中进行训练，训练之后得到分类器模型；

[0012] S6、将学生数据输入分类器模型，输出分类结果。

[0013] 步骤S1中，所述学生数据包含学生一卡通流水记录、学生基本信息、学生成绩和贫困生名单；所述学生基本信息包含学生ID、学生性别、学生名字；学生基本信息包含学生ID、学生性别、学生名字。

[0014] 步骤S1中，所述预处理包含去重、缺失值处理和格式化；

[0015] 所述去重为：将学生数据按学生ID进行排序，通过比较邻近记录是否相似来检测记录是否重复，重复则删除重复记录；

[0016] 所述缺失值处理为：学生数据中某个记录的某个字段为空，则使用平均值进行填充；

[0017] 所述格式化为：将消费时间格式化为yyyy-MM-dd；消费金额统一单位为分，超限则四舍五入；通过预处理，是数据更合理。

[0018] 步骤S2具体过程为：

[0019] U1、从学生一卡通流水记录构造特征；从时间维度、地点维度和交易维度统计均值

和方差；

[0020] U2、将学生一卡通流水记录和学生基本信息、学生成绩、贫困生名单，进行归一化数据特征；

[0021] U3、使用随机森林对特征重要性进行排名，根据排名，选择前30个特征。

[0022] 使用随机森林对特征重要性进行排名，具体为：

[0023] Y1、设定N个样本，每个样本有M个特征；

[0024] Y2、从N个样本中有放回的随机抽取，抽取N次，作为训练一棵决策树的样本；

[0025] Y3、每个节点随机抽取m个特征，m＜M，从中选取信息增益最大的特征作为决策树的分裂节点，在决策树成长的过程中，m值保持不变；

[0026] Y4、重复步骤Y2、Y3，建立大量的决策树，构成随机森林；

[0027] Y5、计算每个特征在随机森林中每棵树上的评分均值，作为特征重要性依据。

[0028] 步骤S3，具体过程为：

[0029] 基于对局部核函数和全局核函数，构造混合核函数，并植入支持向量机SVM 中：

[0030] $$K(x_i,x_j)=(1-\rho)\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)+\rho\left[(x_i\cdot x_j)+c\right]^d,0<\rho<1,$$

[0031] 其中，ρ为混合核函数权系数；

[0032] 步骤S4中，所述寻优过程具体如下：

[0033] V1、设置参数：初始种群数量为60，选择代购为0.8，交叉概率为0.6，变异概率为0.06；

[0034] V2、使用遗传算法确认混合核函数最优混合核函数参数，确认惩罚因子和确认混合核函数权系数；

[0035] V3、混合核函数参数、混合核函数权系数和惩罚因子采用二进制编码，并把其二进制编码组合得到个体染色体基因串，构造出多个染色体组合一个初始种群；

[0036] V4、根据初始种群计算适应度值：

[0037] $$P=\frac{TP}{TP+FP},$$

[0038] $$R=\frac{TP}{TP+FN},$$

[0039] 其中，P为查准率，R为查全率，TP为真正例数目，FP为假正例数目，FN 为假反例数目；

[0040] ρ决定了核函数在混合核函数中的比重；若ρ>0.5，则全局核函数占主导；若ρ<0.5，局部核函数占主导；否则二者重要程度相当。可通过调节ρ来灵活组合局部核函数和全局核函数，同时发挥二者长处。

[0041] 设遗传算法中的适应度值为f(Xi)，即10折交叉验证的macroF1值，则有：

[0042] $$\text{macro}P=\frac{1}{10}\sum_{i=1}^{10}P_i,$$

[0043] $$\text{macro}R=\frac{1}{10}\sum_{i=1}^{10}R_i,$$

[0044] $$\text{macroF1}=\frac{2*macroP*macroR}{macroP+macroR},$$

[0045] 其中，P_i为第i次训练查准率；macroP为宏查准率，是10次训练查准率平均值；R_i为第i次训练查全率；macroR为宏查全率，是10次训练查全率平均值；macroF1为宏F1，是基于宏查准率和宏查全率的调和平均值，即为适应度值；

[0046] V5、根据适应度值计算染色体入选种群概率：

[0047] $$p(X_i)=\frac{f(X_i)}{\sum f(X_j)},$$

[0048] 其中，$p(X_i)$为第i个染色体入选种群概率，X_i为第i个染色体；

[0049] V6、根据入选种群概率的高低，选择代沟为0.8，即保留概率较高的80％染色体，将保留的染色体进行交叉运算和变异运算：

[0050] 所述交叉运算为随机选取两条染色体，随机选择一个交配点做单点杂交，将产生的新的两条染色体代替原来的染色体，放回初始种群；交叉运算概率为 0.6；

[0051] 所述变异运算为杂交后的个体进行变异运算，随机选取一条染色体；

[0052] V7、通过不断进化，获取最优混合核函数系数、最优确认惩罚因子和最优确认混合核函数权系数。

[0053] 步骤S5，具体过程为：

[0054] 根据步骤S4获得的最优混合核函数系数，使用SMO算法通过训练学生数据得到最优的$\hat{a}$、$\hat{b}$，其中，$\hat{a}$为拉格朗日乘子的最优解，$\hat{b}$为分类超平面的最优解；即：SMO每次选取两个拉格朗日乘子，固定其余参数；求解：

[0055] $$\min L(a)=\frac{1}{2}\sum_{i,j=1}^{l}a_i a_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{l}a_i,$$

[0056] $$s.t\ \sum_{i=1}^{l}a_i y_i=0, 0\le a_i\le C, i=1,2,\ldots,l,$$

[0057] 其中，a_i、a_j为拉格朗日乘子；y_i为第i个学生标识，y_j为第j个学生标识；

[0058] 获得更新后的a_i、a_j；

[0059] 求解非线性支持向量机和其对偶问题，重复选取和求解，得到$\hat{a}$、$\hat{b}$；

[0060] 其中非线性支持向量机为：

[0061] $$\begin{cases} min_{\omega,b}\ \frac{1}{2}\omega^2+C\sum_{i=1}^{l}\xi_i \\ s.t\ y_i\left(\omega\cdot\Phi\left(x_i\right)+b\right)\ge 1-\xi_i, i=1,2,\ldots,l \end{cases},$$

[0062] 其中，ω为分类超平面法向量，ξ为松弛变量，$\Phi(x_i)$为将x_i映射后的特征向量；

[0063] 对偶问题：

[0064] $$\begin{cases} \min L(\alpha)=\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{l}\alpha_i \\ s.t\ \sum_{i=1}^{l}\alpha_i y_i=0, 0\le\alpha_i\le C, i=1,2,\ldots,l \end{cases},$$

[0065] 通过$\hat{a}$、$\hat{b}$,得到分类器模型:

[0066] $$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{l}\hat{a}_i y_i K(x_i,x)+\hat{b}\right),$$

[0067] 其中,x为需要认定的学生数据特征值。

[0068] 步骤S6,具体如下:

[0069] 将需要认定的学生数据输入到分类器模型中,通过分类器模型,得出f(x),若为正则表示这个学生大概率为贫困生,若为负则表示这个学生大概率不是贫困生,再通过实际考核,认定新的贫困生,添加到贫困生名单中,得到新的贫困生名单。

[0070] 本发明与现有技术相比,具有如下优点和有益效果:

[0071] 本发明采用遗传算法对混合核函数参数进行寻优,模拟生物的自然选择和遗传机制,用编码空间代替问题参数空间,用适应度作为评价依据,通过随机重组重要基因,让群体中的个体不断进化,逐步接近最优解,减少全局搜索时间,充分发挥局部核函数和全局核函数的优势,在不增加训练时间复杂度的前提下,提高了分类器的推广泛化能力,降低成本。

附图说明

[0072] 图1是本发明一种基于SVM的贫困生辅助认定的方法流程框图;

具体实施方式

[0073] 下面结合实施例及附图对本发明作进一步详细的描述,但本发明的实施方式不限于此。

[0074] 实施例

[0075] 如图1所示,一种基于SVM的贫困生辅助认定方法,包括以下步骤:

[0076] 第一步:采集学生数据,并对学生数据进行预处理;学生数据包含学生一卡通流水记录、学生基本信息、学生成绩和贫困生名单;学生基本信息包含学生ID、学生性别、学生名字;

[0077] 预处理包含去重、缺失值处理和格式化;

[0078] 去重为:将学生数据按学生ID进行排序,通过比较邻近记录是否相似来检测记录是否重复,重复则删除重复记录;

[0079] 缺失值处理为:学生数据中某个记录的摸个字段为空,则使用平均值进行填充;

[0080] 格式化为:将消费时间格式化为yyyy-MM-dd;消费金额统一单位为分,超限则四舍五入。

[0081] 第二步:对预处理学生数据提取特征,通过随机森林对特征重要性进行排名;从一卡通流水记录中构造特征,即各时间段、各地点的消费、充值等行为的总额、均值计数等统

计量。其中，时间维度可分为一天、周末、早、中、晚等几个时间段，地点维度可分为饭堂、商铺、图书馆、西餐厅，交易维度分为消费和充值，对交易金额的统计量分为均值、方差、计数等。比如学生周末在图书馆的消费总额、早上八点前在饭堂的消费均值、在商铺西餐厅的消费次数和均值等；具体过程为：

[0082] 从学生一卡通流水记录构造特征；从时间维度、地点维度和交易维度统计均值和方差；

[0083] 将学生一卡通流水记录和学生基本信息、学生成绩、贫困生名单，进行归一化数据特征；

[0084] 使用随机森林对特征重要性进行排名，根据排名，选择前30个特征。

[0085] 第三步：在满足Mercer条件下，构造混合核函数，并植入支持向量机SVM 中；基于对局部核函数和全局核函数，构造混合核函数：

[0086] $$K(x_i,x_j)=(1-\rho)\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)+\rho\left[(x_i\cdot x_j)+c\right]^d,0<\rho<1,$$

[0087] 其中，ρ为混合核函数权系数，$\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)$为高斯核，属于局部核函数；σ为高斯核的带宽，σ＞0，$[(x_i\cdot x_j)+c]^d$为多项式核，属于全局核函数，c为自由参数，c≥0；d为多项式次数，d≥1，x_i为第i个样本的特征值向量，x_j为第j个特征值向量；将混合核函数植入SVM中。

[0088] 第四步：使用遗传算法对混合核函数参数进行寻优，得到最优混合核函数参数；寻优过程具体如下：

[0089] 设置参数：初始种群数量为60，选择代购为0.8，交叉概率为0.6，变异概率为0.06；

[0090] 使用遗传算法确认混合核函数最优混合核函数参数，确认惩罚因子和确认混合核函数权系数；

[0091] 混合核函数参数(σ、c、d)、混合核函数权系数ρ和惩罚因子C采用二进制编码，并把其二进制编码组合得到个体染色体基因串，群体空间如下：

[0092]

c_1	……	c_{n_1}	σ_1	……	σ_{n_2}	c_1	……	c_{n_3}	d_1	……	d_{n_3}	ρ_1	……	ρ_{n_4}

[0093] 假设$n_1=n_2=n_3=n_4=7$，则这五个参数的二进制编码都是七位，每一位取值0或1，则每个参数的取值范围是0～127。比如：

[0094]

1	……	1	……	1	……	1	……	1	……	1	……	1	……	1

[0095] 其中一个染色体，表示五个参数都是127。

[0096]

0	……	0	……	0	……	0	……	0	……	0	……	0	……	0

[0097] 另外一个染色体，表示五个参数都是0。以此类推，可以构造出多个染色体构成一个初始种群。而后，根据这个初始种群计算适应度值。

[0098] 用个体染色体基因串，构造出多个染色体组合一个初始种群；

[0099] 根据初始种群计算适应度值：

[0100] $$P=\frac{TP}{TP+FP},$$

[0101] $$R=\frac{TP}{TP+FN},$$

[0102] 其中，P为查准率，R为查全率，TP为真正例数目，FP为假正例数目，FN 为假反例数目；

[0103] 设遗传算法中的适应度值为$f(X_i)$，即10折交叉验证的macroF1值，则有：

[0104] $$\mathrm{macro}P=\frac{1}{10}\sum_{i=1}^{10}P_i,$$

[0105] $$\mathrm{macro}R=\frac{1}{10}\sum_{i=1}^{10}R_i,$$

[0106] $$\mathrm{macroF1}=\frac{2*macroP*macroR}{macroP+macroR},$$

[0107] 其中，P_i为第i次训练查准率；macroP为宏查准率，是10次训练查准率平均值；R_i为第i次训练查全率；macroR为宏查全率，是10次训练查全率平均值； macroF1为宏F1，是基于宏查准率和宏查全率的调和平均值，即为适应度值；

[0108] 根据适应度值计算染色体入选种群概率：

[0109] $$\mathrm{p}(X_i)=\frac{f(X_i)}{\sum f(X_j)},$$

[0110] 其中，$p(X_i)$为第i个染色体入选种群概率，X_i为第i个染色体，$f(X_j)$为第j 个染色体的适应度值；

[0111] 根据入选种群概率的高低，选择代沟为0.8，即保留概率较高的80％染色体，将保留的染色体进行交叉运算和变异运算：

[0112] 交叉运算为随机选取两条染色体，随机选择一个交配点做单点杂交，将产生的新的两条染色体代替原来的染色体，放回初始种群；交叉运算概率为0.6；

[0113]

1	……	1	……	1	……	1	……	1	……	1	……	1	……	1
0	……	0	……	0	……	0	……	0	……	0	……	0	……	0

[0114] 单点杂交后：

[0115]

0	……	0	……	0	……	1	……	1	……	1	……	1	……	1
1	……	1	……	1	……	0	……	0	……	0	……	0	……	0

[0116] 变异运算为杂交后的个体进行变异运算，随机选取一条染色体；

[0117]

0	……	0	……	0	……	0	……	0	……	0	……	0	……	0

[0118] 变异运算后：

[0119]

0	……	0	……	0	……	1	……	1	……	1	……	1	……	1

[0120] 遗传算法模拟生物的自然选择和遗传机制，用编码空间代替问题的参数空间，用

适应度函数作为评价依据。通过随机重组重要的基因，让群体中的个体不断进化，逐步接近最优解，并减少全局搜索时间。

[0121] 通过不断进化，获取最优混合核函数系数、最优确认惩罚因子和最优确认混合核函数权系数，即得到多项式核函数与径向基核函数的调整比重，混合核函数的权系数ρ＝0.8253，以及C＝5.9801、σ＝0.0192、c＝0、d＝2。

[0122] 第五步：将最优混合函数系数代入最优分类函数，并对学生数据进行训练，得到分类器模型；具体过程为：

[0123] 根据步骤S4获得的最优混合核函数系数，使用SMO算法通过训练学生数据得到最优的$\hat{a}$、$\hat{b}$，其中，$\hat{a}$为拉格朗日乘子的最优解，$\hat{b}$为分类超平面的最优解；即：SMO每次选取两个拉格朗日乘子，固定其余参数；求解：

[0124] $$\min L(a)=\frac{1}{2}\sum_{i,j=1}^{l}a_i a_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{l}a_i ,$$

[0125] $$s.t\ \sum_{i=1}^{l}a_i y_i=0,0\le a_i\le C,i=1,2,\ldots,l ,$$

[0126] 其中，a_i、a_j为拉格朗日乘子；y_i为第i个学生标识，y_j为第j个学生标识；

[0127] 获得更新后的a_i、a_j；

[0128] 求解非线性支持向量机和其对偶问题，重复选取和求解，得到$\hat{a}$、$\hat{b}$；

[0129] 其中非线性支持向量机为：

[0130] $$\begin{cases} min_{\omega,b}\ \frac{1}{2}\omega^2+C\sum_{i=1}^{l}\xi_i \\ s.t\ y_i\left(\omega\cdot\Phi\left(x_i\right)+b\right)\ge 1-\xi_i,i=1,2,\ldots,l \end{cases},$$

[0131] 对偶问题：

[0132] $$\begin{cases} \min L\left(\alpha\right)=\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(x_i,x_j\right)-\sum_{i=1}^{l}\alpha_i \\ s.t\ \sum_{i=1}^{l}\alpha_i y_i=0,0\le\alpha_i\le C,i=1,2,\ldots,l \end{cases},$$

[0133] 通过$\hat{a}$、$\hat{b}$，得到分类器模型：

[0134] $$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{l}\hat{a}_i y_i K(x_i,x)+\hat{b}\right),$$

[0135] 其中，x为需要认定的学生数据特征值。

[0136] 第六步：将需要认定的学生数据输入到分类器模型中，通过分类器模型计算，得出f(x)，若f(x)为正则表示这个学生大概率为贫困生，若f(x)为负则表示这个学生大概率不是贫困生，再通过实际考核，认定新的贫困生，添加到贫困生名单中，得到新的贫困生名单。

[0137] 上述实施例为本发明较佳的实施方式，但本发明的实施方式并不受上述实施例的限制，其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化，

均应为等效的置换方式，都包含在本发明的保护范围之内。

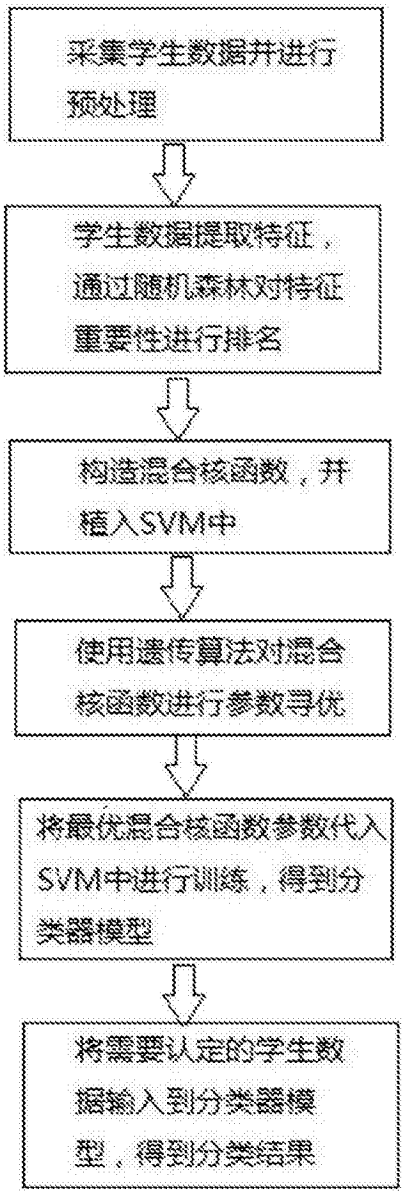

采集学生数据并进行预处理
学生数据提取特征，通过随机森林对特征重要性进行排名
构造混合核函数，并植入SVM中
使用遗传算法对混合核函数进行参数寻优
将最优混合核函数参数代入SVM中进行训练，得到分类器模型
将需要认定的学生数据输入到分类器模型，得到分类结果

图1