(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 107274020 A
(43)申请公布日 2017.10.20

(21)申请号 201710454618.9

(22)申请日 2017.06.15

(71)申请人 北京师范大学
地址 100875 北京市海淀区新街口外大街19号

(72)发明人 余胜泉 卢宇 杨博达 李葆萍

(74)专利代理机构 北京科迪生专利代理有限责任公司 11251
代理人 杨学明 顾炜

(51)Int.Cl.
G06Q 10/04(2012.01)
G06Q 50/20(2012.01)
G06F 17/30(2006.01)

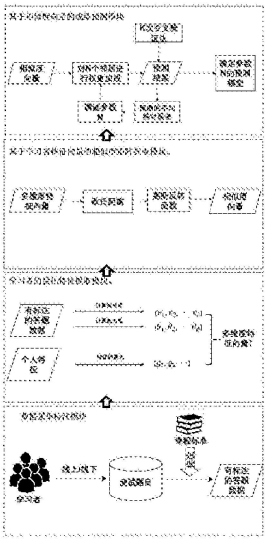
权利要求书3页 说明书5页 附图2页

(54)发明名称

一种基于协同过滤思想的学习者学科总测成绩预测系统及方法

(57)摘要

本发明涉及一种基于协同过滤思想的学习者学科总测成绩预测系统及方法，包括：数据采集标注模块、学习者的量化特征提取模块、基于学习者量化特征的相似度向量提取模块、基于相似度向量的成绩预测模块。本发明可以解决对学习者学业成绩的预测问题，适用于一般在线学习平台和系统，也可以应用于实际教学评估和诊断中，为学习者提供个性化的教学服务，提高学习针对性和学习效率。



CN 107274020 A

1.一种基于协同过滤思想的学习者学科总测成绩预测系统，其特征在于：包括数据采集标注模块、学习者的量化特征提取模块、基于学习者量化特征的相似度向量提取模块和基于相似度向量的成绩预测模块；其中：

数据采集标注模块：根据具体学科科目的课程标准，对该学科科目的知识点进行划分，按照学习的时间顺序排列划分知识点；学习者在对每个知识点进行学习后，进行该知识点的水平测试，测试将通过线上电子化课堂或者线下课堂或作业的形式进行，从而收集各知识点对应的测试数据和成绩。测试数据中包括题目本身及题目标注所属的知识点，每一个知识点都包含至少一道以上的测试题目，每个知识点所包含的测试题目数量可以不等；同时，在测试过程中，收集学习者本身的基础数据，包括所在学校及地区；

学习者的量化特征提取模块：基于数据采集标注模块中所收集的基础数据，计算学习者i在知识点p的得分率v_p：

v_p＝学习者答对的p的题目的数量/p涵盖的题目的数量

对每个知识点计算得分率，得到学习者i在每个知识点的能力值$V_i=\{v_p|p\in P\}$，此处的P为某一学习过程中知识点p的集合；除此之外，根据项目反映理论，通过整合学习者i对于每个知识点答题情况，得到学习者i在每个知识点的能力值$\emptyset_i=\{\theta_p|p\in P\}$，完成所有知识点P的学习者i的成绩测试层面，该学习者对应的特征向量有得分率向量$V_i=\{v_p|p\in P\}$和能力值向量$\emptyset_i=\{\theta_p|p\in P\}$；同时，将学习者i的所在的学校、地区的基础数据进行量化，作为补充特征向量G_i，来细化学习者个体区别之间的差异，从而形成多维度特征向量；最终，学习者i的多维度特征向量$T_i=[V_i,\emptyset_i,G_i]$，包括已学习过的知识点的得分率向量$V_i$，能力值向量$\emptyset_i$以及学习者个体特征向量$G_i$；

基于学习者多维度特征向量的相似度计算模块：根据学习者的量化特征提取模块产生的学习者i的多维度特征T_i，计算T_i与具有相同学习过程的其他学习者j的欧氏距离，从而得到学习者i与其余学习者之间的距离向量$\{D_{ij}|j\in J\}$，其中J为其余学习者的集合，然后利用高斯函数作为反转函数将学习者i与其余学习者j之间的欧氏距离D_{ij}，转变为学习者i与其余学习者j之间的相似度S_{ij}；

基于相似度向量的成绩预测模块：基于学习者多维度特征向量的相似度计算模块中得到的学习者i与其余学习者J之间的相似度向量$\{S_{ij}|j\in J\}$。从J个其余学习者中，筛选出前N个与学习者i相似度最高的学习者，J_N表示这N个学习者的集合。以学习者i与挑选出的N个学习者的相似度$\{S_{ij}|j\in J_N\}$作为权重，用N个学习者学业成绩$\{Y_j|j\in J_N\}$进行加权平均，从而预测学习者i的成绩。预测的准确率随着N的变化而变化，在进行预测前需要先根据预测效果调试N的数值。

2.根据权利要求1所述的一种基于协同过滤思想的学习者学科总测成绩预测系统，其特征在于：所述学习者的量化特征提取模块中，利用项目反应理论计算学习者i在每个知识点的能力值$\emptyset_i=\{\theta_p|p\in P\}$，具体方法如下：

在测试数据中，任一知识点p往往包含多个题目，知识点p下的题目表示为$\{k|k\in p\}$，学习者i在知识点p下的答题表现$R_p^i=\{r_k^i|k\in p\}$，其中r_k^i表示学习者i对题目k的作答结果，当作答结果正确时r_k^i=1；当作答结果错误时，$r_k^i=0$。基于项目反映理论，学习者i的能力值跟

其答对题目k的概率满足下方的双参数模型：

$$f(\theta_i)=\frac{1}{1+e^{-a_k(\theta_i-b_k)}}$$

其中θ_i表示学习者i在知识点p的能力，参数a_k与b_k分别为题目k的区分度与难度，$f(\theta_i)$为学习者正确作答该题目的概率；

已知所有学习者在知识点p下的答题表现$\{R_p^i|i\in M\}$，此处M为所有学习者的集合，通过使用最大期望算法来寻得到每个学习者对于知识点p的能力$\{\theta_p^i|i\in M\}$和每道题目的难度$\{b_k|k\in p\}$和区分度$\{a_k|k\in p\}$，目标似然函数数学表达为$\prod_{i\in M}\prod_{k\in p}f_k(\theta_i)^{r_k^i}(1-f_k(\theta_i))^{1-r_k^i}$。

3.根据权利要求1所述的一种基于协同过滤思想的学习者学科总测成绩预测系统，其特征在于：所述基于学习者多维度特征向量的相似度计算模块中，利用高斯函数作为反转函数将学习者i与其余学习者j之间的欧氏距离D_{ij}，转变为学习者i与其余学习者j之间的相似度S_{ij}，具体实现如下：

$$S_{ij}=f(D_{ij};\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(D_{ij}-\mu)^2}{2\sigma^2}\right)$$

其中的S_{ij}为学习者i与学习者j之间的相似度，D_{ij}为学习者i与学习者j的综合特征向量间的欧氏距离，σ和μ为决定高斯函数形状的常数。

4.根据权利要求1所述的一种基于协同过滤思想的学习者学科总测成绩预测系统，其特征在于：所述基于相似度向量的成绩预测模块中，调试参数N的方法如下：

利用平均绝对误差(MAE)作为主要调参依据，其中$MAE=\sum_{i=1}^{n}|\hat{y}_i-y_i|/n$，n表示对n个学习者做了学业成绩的预测，$\hat{y}_i$表示预测的学习者i的学业成绩，$y_i$表示学习者i的实际成绩；具体调整过程中利用K次交叉验证方法计算得出的K个平均绝对误差(MAE)的平均值来作为平均绝对误差，选取使平均绝对误差最小的N值来作为模型参数。

5.一种基于协同过滤思想的学习者学科总测成绩预测方法，其特征在于步骤如下：

(1)数据采集标注：根据具体学科科目的课程标准，对该学科科目的知识点进行划分，按照学习的时间顺序排列划分知识点；学习者在对每个知识点进行学习后，进行该知识点的水平测试，测试将通过线上电子化课堂或者线下课堂或作业的形式进行，从而收集各知识点对应的测试数据和成绩；测试数据中包括题目本身及题目标注所属的知识点，每一个知识点都包含至少一道以上的测试题目，每个知识点所包含的测试题目数量可以不等；同时，在测试过程中，收集学习者本身的基础数据，包括所在学校及地区；

(2)学习者的量化特征提取：基于数据采集标注模块中所收集的基础数据，计算学习者i在知识点p的得分率v_p：

v_p＝学习者答对的p的题目的数量/p涵盖的题目的数量

对每个知识点计算得分率，得到学习者i在每个知识点的能力值$V_i=\{v_p|p\in P\}$，此处的P为某一学习过程中知识点p的集合；除此之外，根据项目反映理论，通过整合学习者i对于每个知识点答题情况，得到学习者i在每个知识点的能力值$\emptyset_i=\{\theta_p|p\in P\}$，完成所有知识点P的学习者i的成绩测试层面，该学习者对应的特征向量有得分率向量$V_i=\{v_p|p\in P\}$和能力值向量$\emptyset_i=\{\theta_p|p\in P\}$；同时，将学习者i的所在的学校、地区的基础数据进行量化，作为

补充特征向量G_i，来细化学习者个体区别之间的差异，从而形成多维度特征向量；最终，学习者i的多维度特征向量$T_i = [V_i, \emptyset_i, G_i]$，包括已学习过的知识点的得分率向量$V_i$，能力值向量$\emptyset_i$以及学习者个体特征向量$G_i$；

(3)基于学习者多维度特征向量的相似度计算：根据学习者的量化特征提取模块产生的学习者i的多维度特征T_i，计算T_i与具有相同学习过程的其他学习者j的欧氏距离，从而得到学习者i与其余学习者之间的距离向量$\{D_{ij} | j \in J\}$，其中J为其余学习者的集合，然后利用高斯函数作为反转函数将学习者i与其余学习者j之间的欧氏距离D_{ij}，转变为学习者i与其余学习者j之间的相似度S_{ij}；

(4)基于相似度向量的成绩预测：基于学习者多维度特征向量的相似度计算模块中得到的学习者i与其余学习者J之间的相似度向量$\{S_{ij} | j \in J\}$。从J个其余学习者中，筛选出前N个与学习者i相似度最高的学习者，J_N表示这N个学习者的集合。以学习者i与挑选出的N个学习者的相似度$\{S_{ij} | j \in J_N\}$作为权重，用N个学习者学业成绩$\{Y_j | j \in J_N\}$进行加权平均，从而预测学习者i的成绩。预测的准确率随着N的变化而变化，在进行预测前需要先根据预测效果调试N的数值。

一种基于协同过滤思想的学习者学科总测成绩预测系统及方法

技术领域

[0001] 本发明涉及一种基于协同过滤思想的学习者学科总测成绩预测系统及方法，属于数据挖掘技术，特别是涉及教育领域的数据挖掘。

背景技术

[0002] 数据挖掘是一种基于大量数据进行信息提取和知识发现的方法，数据挖掘中的一些方法包括聚类、关联规则学习、相关性分析、回归性分析以及分类等已经被广泛应用于互联网、工业制造、交通等各个领域。其中一类基于协同过滤思想的数据挖掘算法可以有效筛选出相似群体，故该算法已经成熟应用于电商推荐系统上来寻找相似兴趣品味的用户并进行推荐。在教育领域，此类算法的应用相对比较新颖，而且在教育技术领域有很大的应用前景。本发明首次提出将该算法用于学习者学科总测学习预测的问题上。做到了提前预测学习者对于未来的知识的学习效果。该方法的实现可以用来支持教育决策、对学习者进行信息和课程内容的推荐、学习者学习过程中的提前预警、学习者专业选择推荐以及制定学习者个性化的学习策略等。

发明内容

[0003] 本发明要解决的问题是：克服现有技术的不足，将教育学测量手段跟数据挖掘技术相结合，提供一种基于协同过滤思想的学科总测成绩预测系统及方法，对学习者知识点和整体学科的掌握状态进行预测和估计，从而为学习者提供个性化的教学服务，提高学习针对性和学习效率。

[0004] 本发明解决其问题所采用的方案是：一种基于协同过滤思想的学习者学科总测成绩预测系统，包括数据采集标注模块、学习者的量化特征提取模块、基于学习者量化特征的相似度向量提取模块、基于相似度向量的成绩预测模块，其中：

[0005] 数据采集标注模块：根据具体学科科目的课程标准，系统对该学科的知识点进行划分，按照时间顺序排列划分后的知识点。学习者在对每个知识点进行学习后，进行该知识点的水平测试。测试将通过线上电子化课堂或者线下课堂或作业的形式进行，从而收集各知识点对应的测试数据和成绩。测试数据中包括题目本身及题目标注所属知识点。每一个知识点都包含至少一道以上的测试题目，每个知识点对应的测试题目数量可以不等。同时，在测试过程中，收集学习者本身的个体基础数据，例如所在地区和学校。

[0006] 学习者的量化特征提取模块：基于模块一中所收集的数据，可以分别针对每个学习者，计算其在知识点p的得分率v_p：

[0007] v_p＝p下答对的题目的数量/p下包含的所有题目的数量。

[0008] 因此，对于完成P个知识点的个体学习者i的成绩测试层面，该学习者对应的基本特征向量$V_i=\{v_p|p\in P\}$。除此之外，根据项目反映理论，该系统还可以通过整合学习者i对于每个知识点答题情况，得到学习者i在每个知识点的能力值$\emptyset_i=\{\theta_p|p\in P\}$。因此，对于完

成P个知识点的学习者i的成绩测试层面，该学习者对应的特征向量有得分率向量$V_i=\{v_p|p\in P\}$和能力值向量$\emptyset_i=\{\theta_p|p\in P\}$。同时，将学习者i的所在的地区、学校等个体特征进行量化，作为补充特征向量G_i，来细化学习者个体区别之间的差异，从而形成多维度特征向量。具体来说，学习者i的多维度特征向量$T_i=[V_i,\emptyset_i,G_i]$，其包括已学习过的知识点的得分率向量$V_i$，能力值向量$\emptyset_i$以及学习者个体特征向量$G_i$。

[0009] 基于学习者多维度特征向量的相似度计算模块：基于学习者的量化特征提取模块产生的学习者i的多维度特征T_i，计算T_i与系统中具有相同学习过程的其他学习者的多维度特征T_j的欧式距离。从而得到学习者i与其余学习者之间的欧氏距离向量$\{D_{ij}|j\in J\}$，其中J为其余学习者的集合。为了进一步得到相似度的数值，需要利用反转函数将学习者i与学习者j之间的欧氏距离D_{ij}转化为相似度S_{ij}。此模块中使用高斯函数作为反转函数，将学习者i与其余学习者j之间的欧氏距离向量$\{D_{ij}|j\in J\}$，转变为学习者i与其余学习者j之间的相似度向量$\{S_{ij}|j\in J\}$。

[0010] 基于相似度向量的成绩预测模块：给定系统中学习者i的学科总测成绩Y_i是待预测的；系统中储存的历史数据包含的其余学习者J的学科总测成绩$\{Y_j|j\in J\}$是已知。根据得到的学习者i与其余学习者J之间的相似度向量$\{S_{ij}|j\in J\}$，本模块从J个其余学习者中，筛选出前N个与学习者i相似度最高学习者。此处用J_N表示这N个学习者的集合。以学习者i与挑选出的N个学习者的相似度$\{S_{ij}|j\in J_N\}$作为权重，用N个相似度高的学习者学业成绩$\{Y_j|j\in J_N\}$进行加权平均，最终预测学习者i的总测成绩$\hat{Y}_i$。由于系统预测的准确率随着N的变化而变化，故在进行预测前需要先根据系统预测效果调试N的数值。

[0011] 系统参数的调试方法：

[0012] 由于系统预测的准确率随着N的变化而变化，故在基于相似度向量的成绩预测模块中，需要对算法中的参数N进行调试，得到合适的N的数值，最终得到可以最准确预测的系统模型。具体调试参数N的方法如下：

[0013] 1)给N一个初始值，以一个常数递增，分别计算不同的N下，系统的预测效果。一般情况下随着N的增加，系统的预测误差先减小，后增加。故当随着N的增加，系统的预测误差不再减小时，那么此时的N就被选取为系统中最终的常数N。

[0014] 2)模型预测的误差大小的评判标准为平均绝对误差(MAE)为：$MAE=\sum_{i=1}^{n}|\hat{y}_i-y_i|/n$。n表示系统对n个学习者做了学业成绩的预测。$\hat{y}_i$表示系统预测的学习者i的学业成绩。$y_i$表示学习者i的实际成绩。

[0015] 3)对于某一给定N值的系统。根据系统中已经储存的学习者，使用K次交叉验证法计算得出的K个平均绝对误差(MAE)的平均值来作为系统的平均绝对误差。通过变化N值，当系统的平均绝对误差不再减小时，那么此时的N就被选取为系统中最终的常数N。

[0016] 本发明与现有方法相比的有益效果为：

[0017] (1)本发明可以解决对学习者科目总测成绩的预测的问题，为学习者提前预警，提高了学习针对性和学习效率。

[0018] (2)本发明方法将数据挖掘技术和教育测量手段相结合。针对学习者科目总测成绩的预测问题，提出了利用测试数据结合学习者能力和个人特征数据提取出多维度的特征向量。然后，基于协同过滤思想，建立预测模型，最终给出学习者总测成绩的预测结果。

附图说明

[0019] 图1为本发明一种基于协同过滤思想的学科总测成绩预测系统的结构图；

[0020] 图2为本发明的学习者能力特征提取流程；

[0021] 图3为本发明的中使用的交叉验证流程；

[0022] 图4为本发明系统中的数据储存结构。

具体实施方式

[0023] 下面结合附图及具体实施方式详细介绍本发明。

[0024] 如图1所示，本发明为一种基于协同过滤思想的学习者学科总测成绩预测系统，包括：数据采集标注模块、学习者的量化特征提取模块、基于学习者量化特征的相似度向量提取模块、基于相似度向量的成绩预测模块。

[0025] 数据采集标注模块具体实现如下：

[0026] 根据具体学科科目的课程标准，系统对该学科的知识点进行划分，按照时间顺序排列划分后的知识点。例如：数学学科某一年级的知识点划分和知识点的学习时间顺序如下：有理数→一元一次方程→几何体→线段→角→相交线→平行线。学习者在对每个知识点进行学习后，进行该知识点的水平测试。测试将通过线上电子化课堂或者线下课堂或作业的形式进行，从而收集各知识点对应的测试数据和成绩。测试数据中包括题目本身及题目标注所属知识点。每一个知识点都包含至少一道以上的测试题目，每个知识点对应的测试题目数量可以不等。同时，在测试过程中，收集学习者本身的个体基础数据，例如所在地区和学校。学习者应涵盖同一年纪各层次水平的人群。例如，可以是某一地区同一年级所有的在籍学生；对于每个学科，训练数据的规模应保持在一定规模以上。例如3000个学习者对于数学学科14个知识点的独立测试结果。数据将以图4的结构储存：每个知识点对应一张表，表中每一行对应一名学习者在该知识点下各个题目上的测试结果。

[0027] 基于采集标注的信息进行学习者能力特征提取，具体实现如下：

[0028] 基于模块一中所收集的数据，可以分别对每个学习者，计算其在知识点p的得分率v_p：

[0029] v_p＝p下答对的题目的数量/p下包含的所有题目的数量。

[0030] 例如某学习者在一元一次不等式这个学科答对了5道题，该学科下一共包含了10道题，则学习者在该学科下的得分率v＝0.5。对于完成P个知识点的个体学习者i的成绩测试层面，该学习者对应的基本特征向量$V_i=\{v_p|p\in P\}$。除此之外，根据项目反映理论，该系统还可以通过整合所有学习者i对于每个知识点答题情况，得到学习者i在每个知识点的能力值$\emptyset_i=\{\theta_p|p\in P\}$。如图2所示，例如某学习者在一元一次不等式这个学科下的十道题目的答题情况如下[1,0,1,1,1,0,0,0,1,0]。向量里1表示回答正确，0表示回答错误。根据学习者的答题情况，寻找到的使目标似然函数最大的能力值，便是该学习者的能力值θ。因此，对于完成P个知识点的学习者i的成绩测试层面，该学习者对应的特征向量有得分率向量$V_i=\{v_p|p\in P\}$和能力值向量$\emptyset_i=\{\theta_p|p\in P\}$。同时，将学习者i的所在的地区、学校等个体特征进行量化，作为补充特征向量G_i，来细化学习者个体区别之间的差异，从而形成多维度特

征向量。例如某一学习者所在的地区的数学平均分为72，其所在学校的数学平均分为69，那么该学习者补充特征向量G＝[72,69]。最后，学习者i的多维度特征向量$T_i=[V_i,\emptyset_i,G_i]$，其包括已学习过的知识点的得分率向量$V_i$，能力值向量$\emptyset_i$以及学习者个体特征向量$G_i$。

[0031] 基于学习者多维度特征向量的相似度计算模块，其具体实现如下：

[0032] 基于学习者的量化特征提取模块产生的学习者i的多维度特征T_i，计算T_i与系统中具有相同学习过程的其他学习者的多维度特征T_j的欧式距离。从而得到学习者i与其余学习者之间的欧氏距离向量$\{D_{ij}|j\in J\}$，其中J为其余学习者的集合。例如学习者A的多维度向量$T_A=[a_1,a_2,\cdots,a_n]$学习者B的多维度向量$T_B=[b_1,b_2,\cdots,b_n]$。这两个学习者之间的距离$D=\sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2}$。为了进一步得到相似度的数值，需要利用反转函数将学习者i与学习者j之间的欧氏距离D_{ij}转化为相似度S_{ij}。此模块中使用高斯函数作为反转函数，将学习者i与其余学习者j之间的欧氏距离向量$\{D_{ij}|j\in J\}$，转变为学习者i与其余学习者j之间的相似度向量$\{S_{ij}|j\in J\}$。高斯函数的数学表达如下：

[0033] $$S_{ij}=f(D_{ij};\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(D_{ij}-\mu)^2}{2\sigma^2}\right)$$

[0034] 其中的S_{ij}为学习者i与学习者j之间的相似度，D_{ij}为学习者i与学习者j的综合特征向量间的欧氏距离，σ和μ为决定高斯函数形状的常数。通常μ＝0；σ＝1。

[0035] 基于相似度向量的成绩预测模块，其具体实现如下：

[0036] 给定系统中储存的历史数据包含的其余学习者J的学科总测成绩$\{Y_j|j\in J\}$是已知，则对于待预测学科总测成绩Y_i的学习者i，根据得到的学习者i与其余学习者J之间的相似度向量$\{S_{ij}|j\in J\}$，本模块从J个其余学习者中，筛选出前N个与学习者i相似度最高的学习者。此处用J_N表示这N个学习者的集合。以学习者i与筛选出的N个学习者的相似度$\{S_{ij}|j\in J_N\}$作为权重，用N个其余学习者学业成绩$\{Y_j|j\in J_N\}$进行加权平均，最终预测学习者i的总测成绩$\hat{Y}_i$。加权平均的数学方程如下：

[0037] $$\hat{Y}_i=\frac{\sum_{j\in J_N}S_{ij}Y_j}{\sum_{j\in J_N}S_{ij}}$$

[0038] 例如对于学习者A，系统根据其他学习者与A的相似度向量寻找到了前5个跟他相似度最高的学生，那些相似度分别是[1,0.99,0.99,0.83,0.82]。这5个学习者的总测成绩分别是[74,89,83,70,78]，那么根据加权平均，得到学习者A的成绩$\hat{Y}=79.12$。

[0039] 由于不同的N值对系统的预测效果有显著的影响。故需要先调试出合适的N值使系统的预测误差最小。其具体的过程和方法如下：

[0040] (1)一般情况下随着N的增加，系统的预测误差先减小，后增加。给N一个初始值，以一个常数递增，分别计算不同的N下系统的预测效果。例如计算N分别取值5,10,15,20,25时系统的预测误差。当随着N的增加，系统的预测误差不再减小时，那么此时的N就被选取为系统中最终的N值。

[0041] (2)如图3所示，对某一N值。基于系统中已经储存的学习者数，使用K次交叉验证计算得出的K个平均绝对误差(MAE)的平均值来作为系统的平均绝对误差。其中模型预测的误差大小的评判标准为平均绝对误差(MAE)为：$MAE=\sum_{i=1}^{n}|\hat{y}_i-y_i|/n$。n表示系统对n个学习

者做了学业成绩的预测。$\hat{y}_i$表示系统预测的学习者i的学业成绩，y_i表示学习者i的实际成绩。

[0042] (3)当系统的平均绝对误差不再随着N的增加而减小时，此时的数值将作为系统中最终确定的该参数值。

[0043] 本发明未详细阐述的部分属于本领域公知技术。

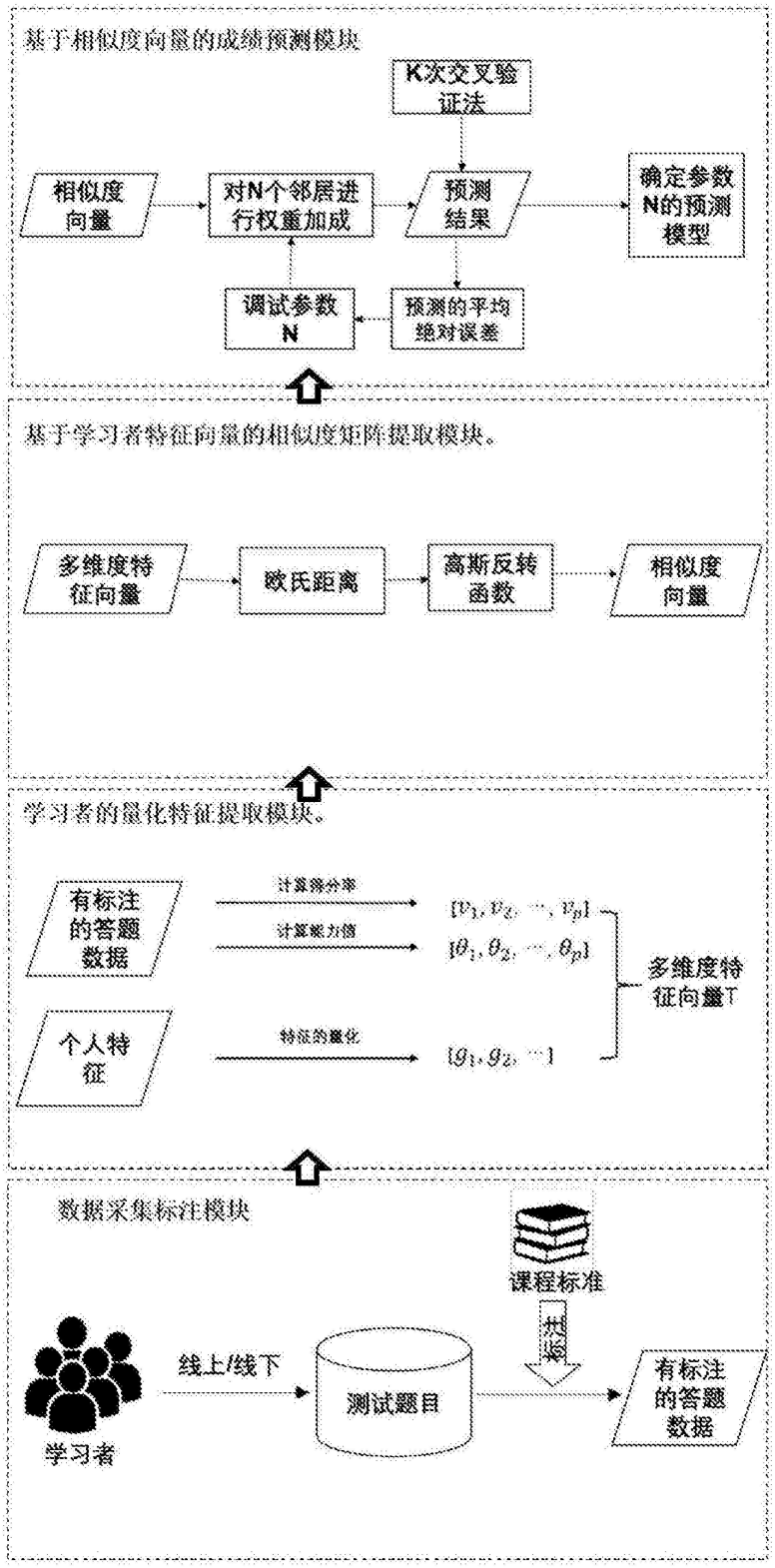

基于相似度向量的成绩预测模块
K次交叉验证法
相似度向量
对N个邻居进行权重加成
预测结果
确定参数N的预测模型
调试参数N
预测的平均绝对误差
基于学习者特征向量的相似度矩阵提取模块。
多维度特征向量
欧氏距离
高斯反转函数
相似度向量
学习者的量化特征提取模块。
有标注的答题数据
计算得分率
计算能力值
多维度特征向量T
个人特征
特征的量化
数据采集标注模块
课程标准
线上/线下
测试题目
有标注的答题数据
学习者

图1

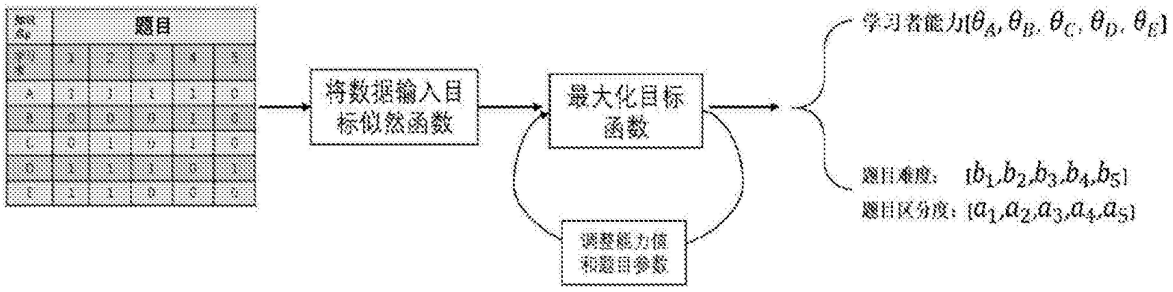


图2

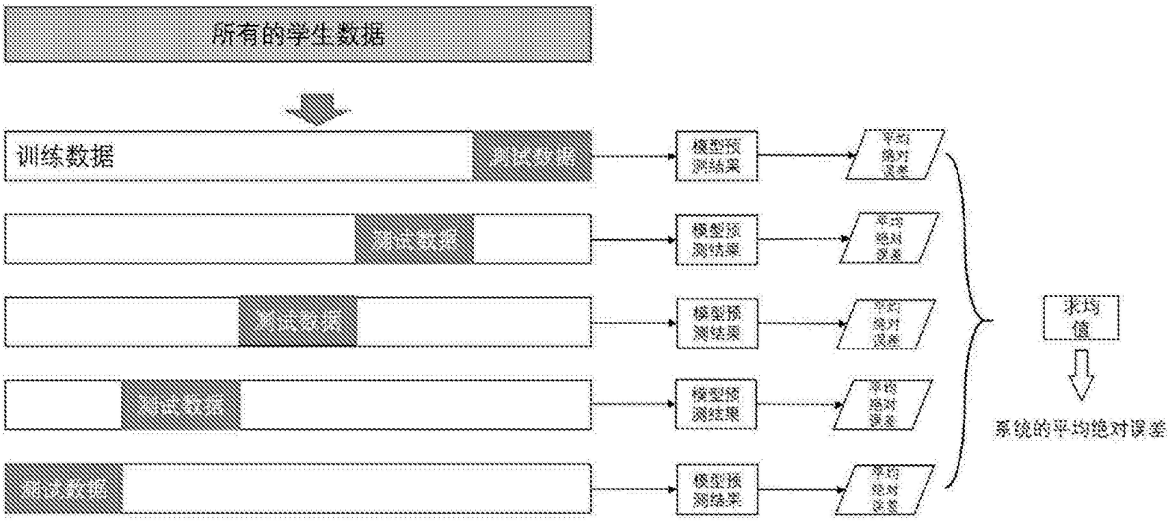


图3

有理数

	题目 1	题目 2	题目 3	……
学习者 001	0	1	1	……
学习者 002	1	1	1	……
学习者 003	1	1	0	……
……	……	……	……	……

一元一次方程

	题目 1	题目 2	题目 3	……
学习者 001	1	1	1	……
学习者 002	1	0	1	……
学习者 003	1	0	0	……
……	……	……	……	……

图4