

课程推荐方法、装置、计算机设备及介质

|  |  |
| --- | --- |
| 申请号： | CN202011590904.6 |
| 申请日： | 20201229 |
| 申请（专利权）人： | [中国平安人寿保险股份有限公司] |
| 地址： | 广东省深圳市福田区益田路5033号平安金融中心14、15、16、37、41、44、45、46层 |
| 发明人： | [黄良仁] |
| 主分类号： | G06F16/9535 |
| 公开（公告）号： | CN112632385B |
| 公开（公告）日： | 20230922 |
| 代理机构： | 深圳市世联合知识产权代理有限公司 |
| 代理人： | [汪琳琳] |

www.patexplorer.com

|  |  |  |
| --- | --- | --- |
| **（19）中华人民共和国国家知识产权局** | | |
|  |  |  |
| **（12）发明专利** | |
| **（10）授权公告号** CN112632385B  **（45）授权公告日** 20230922 | | |

|  |  |
| --- | --- |
| **（21）申请号** CN202011590904.6  **（22）申请日** 20201229  **（71）申请人** [中国平安人寿保险股份有限公司]  **地址** 广东省深圳市福田区益田路5033号平安金融中心14、15、16、37、41、44、45、46层  **（72）发明人** [黄良仁]  **（74）专利代理机构** 深圳市世联合知识产权代理有限公司  **代理人** [汪琳琳] |  |
| **（54）发明名称**  课程推荐方法、装置、计算机设备及介质 |  |
| **（57）摘要**  本发明信息推荐领域，公开了一种课程推荐方法、装置、计算机设备及介质，所述方法包括：获取待推荐用户的基础数据，对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像，根据用户画像进行特征提取，得到目标向量，将目标向量输入至预先训练的课程推荐模型中，并采用预先训练的课程推荐模型进行推荐，得到待推荐课程类别，从待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将目标推荐课程推荐给待推荐用户，本发明不同待推荐用户的个人画像，进行个性化推荐，有利于提高课程推荐的精准程度。 |

|  |
| --- |
| **权 利 要 求 书** |

1.一种课程推荐方法，其特征在于，包括：

获取待推荐用户的基础数据；

对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像；

根据所述用户画像进行特征提取，得到目标向量；

将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别；

从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户；

所述对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像包括：

**对所述基础数据进行数据预处理，得到标准数据；对所述标准数据进行分词处理，并对分词处理结果进行聚类，得到所述待推荐用户的关联类别；根据所述关联类别，生成目标标签和所述目标标签的重要性排序，得到所述用户画像；**

所述根据所述用户画像进行特征提取，得到目标向量包括：

基于预设语料库，构建所述用户画像中每个标签对应的基础词向量；针对每个所述基础词向量，计算该基础词向量与其他基础词向量之间的空间距离，并从所述空间距离中选取最小值作为该基础词向量的最小空间距离；将所述最小空间距离中小于或等于预设空间距离阈值的基础词向量，作为标签向量；基于K-Means聚合算法对所述标签向量进行分类，得到目标向量；

所述预先训练的课程推荐模型为梯度提升决策树模型，所述采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别包括：

将所述目标向量的每个维度作为一个训练特征；将所述训练特征输入到梯度提升决策树模型中，通过梯度提升决策树模型对所述训练特征进行训练，得到n棵决策树，其中，n为正整数；将所每棵所述决策树的路径包含的训练特征作为自变量，基于二分类逻辑回归模型，进行课程类别的预测，得到每个课程类别对应的预测分值；将预测分值超过预设阈值的课程类别，作为所述待推荐课程类别。

2.如权利要求1所述的课程推荐方法，其特征在于，在所述将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别之前，所述课程推荐方法还包括：

获取每个样本用户的用户画像；

提取每个所述用户画像的特征向量；

获取每个所述样本用户的历史行为数据集合；

确定每个所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别；

针对每个所述样本用户，将所述样本用户的用户画像的特征向量作为输入，将所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练得到所述预先训练的课程推荐模型。

3.如权利要求1至2任一项所述的课程推荐方法，其特征在于，在所述从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户之后，所述课程推荐方法还包括：

接收所述待推荐用户针对所述目标推荐课程的学习评估数据；

基于所述学习评估数据，对所述待推荐用户进行二次画像，得到更新画像；

基于所述更新画像，对所述目标推荐课程进行更新。

4.一种课程推荐装置，所述课程推荐装置运行时实现如权利要求1至3任一项所述的课程推荐方法，其特征在于，所述课程推荐装置包括：

采集模块，用于获取待推荐用户的基础数据；

画像模块，用于对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像；

提取模块，用于根据所述用户画像进行特征提取，得到目标向量；

训练模块，用于将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别；

推荐模块，用于从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户。

5.如权利要求4所述的课程推荐装置，其特征在于，所述画像模块包括：

预处理单元，用于对所述基础数据进行数据预处理，得到标准数据；

聚类单元，用于对所述标准数据进行分词处理，并对分词处理结果进行聚类，得到所述待推荐用户的关联类别；

画像生成单元，用于根据所述关联类别，生成目标标签和所述目标标签的重要性排序，得到所述用户画像。

6.一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求1至3任一项所述的课程推荐方法。

7.一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至3任一项所述的课程推荐方法。

|  |
| --- |
| **说 明 书** |

**课程推荐方法、装置、计算机设备及介质**

**技术领域**

本发明涉及信息推荐领域，尤其涉及一种课程推荐方法、装置、计算机设备及介质。

**背景技术**

随着社会的飞速发展和信息技术的进步，各行各业的用户都需要不断学习来提高值的技能，而需要提高哪些技能，如何去进行课程的选择，成了一个焦点。

现有方式中，通常课程推荐是通过设定规则来实现，比如哪些职位需要掌握哪些技能，但不同用户自身情况不同，这种推荐方式针对性较弱，使得推荐效率不高。随着社会的飞速发展和信息技术的进步，各行各业的用户都需要不断学习来提高值的技能，而需要提高哪些技能，如何去进行课程的选择，成了一个焦点。

现有方式中，通常课程推荐是通过设定规则来实现，比如哪些职位需要掌握哪些技能，但不同用户自身情况不同，这种推荐方式针对性较弱，使得推荐精准程度不高。

**发明内容**

本发明实施例提供一种课程推荐方法、装置、计算机设备和存储介质，以提高课程推荐的准确性。

为了解决上述技术问题，本申请实施例提供一种课程推荐方法，包括：

获取待推荐用户的基础数据；

对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像；

根据所述用户画像进行特征提取，得到目标向量；

将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别；

从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户。

可选地，对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像包括：

对所述基础数据进行数据预处理，得到所述标准数据；

对所述标准数据进行分词处理，并对分词处理结果进行聚类，得到所述待推荐用户的关联类别；

根据所述关联类别，生成目标标签和所述目标标签的重要性排序，得到所述用户画像。

可选地，所述根据所述用户画像进行特征提取，得到目标向量包括：

基于预设语料库，构建所述用户画像中每个标签对应的基础词向量；

针对每个所述基础词向量，计算该基础词向量与其他基础词向量之间的空间距离，并从所述空间距离中选取最小值作为该基础词向量的最小空间距离；

将所述最小空间距离中小于或等于预设空间距离阈值的基础词向量，作为标签向量；

基于K-Means聚合算法对所述标签向量进行分类，得到目标向量。

可选地，在所述将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别之前，所述课程推荐方法还包括：

获取每个样本用户的用户画像；

提取每个所述用户画像的特征向量；

获取每个所述样本用户的历史行为数据集合；

确定每个所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别；

针对每个所述样本用户，将所述样本用户的用户画像的特征向量作为输入，将所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练得到所述预先训练的课程推荐模型。

可选地，所述预先训练的课程推荐模型为梯度提升决策树模型，所述采用所述预先训练的课程推荐模型进行训数据处理，得到待推荐课程类别包括：

将所述目标向量的每个维度作为一个训练特征；

将所述训练特征输入到梯度提升决策树模型中，通过梯度提升决策树模型对所述训练特征进行训练，得到n棵决策树，其中，n为正整数；

将所每棵所述决策树的路径包含的训练特征作为自变量，基于二分类逻辑回归模型，进行课程类别的预测，得到每个课程类别对应的预测分值；

将预测分值超过预设阈值的课程类别，作为所述待推荐课程类别。

可选地，在所述从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户之后，所述课程推荐方法还包括：

接收所述待推荐用户针对所述目标推荐课程的学习评估数据；

基于所述学习评估数据，对所述待推荐用户进行二次画像，得到更新画像；

基于所述更新画像，对所述目标推荐课程进行更新。

为了解决上述技术问题，本申请实施例还提供一种课程推荐装置，包括：

采集模块，用于获取待推荐用户的基础数据；

画像模块，用于对所述基础数据进行数据预处理和聚合，根据得到的聚合结果，对所述待推荐用户进行画像处理，得到用户画像；

提取模块，用于根据所述用户画像进行特征提取，得到目标向量；

训练模块，用于将所述目标向量输入至预先训练的课程推荐模型中，并采用所述预先训练的课程推荐模型进行数据处理，得到待推荐课程类别；

推荐模块，用于从所述待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将所述目标推荐课程推荐给所述待推荐用户。

可选地，画像模块包括：

预处理单元，用于对所述基础数据进行数据预处理，得到所述标准数据；

聚类单元，用于对所述标准数据进行分词处理，并对分词处理结果进行聚类，得到所述待推荐用户的关联类别；

画像生成单元，用于根据所述关联类别，生成目标标签和所述目标标签的重要性排序，得到所述用户画像。

可选地，所述提取模块包括：

基础词向量构建单元，用于基于预设语料库，构建所述用户画像中每个标签对应的基础词向量；

空间距离计算单元，用于针对每个所述基础词向量，计算该基础词向量与其他基础词向量之间的空间距离，并从所述空间距离中选取最小值作为该基础词向量的最小空间距离；

词向量筛选单元，用于将所述最小空间距离中小于或等于预设空间距离阈值的基础词向量，作为标签向量；

分类单元，用于基于K-Means聚合算法对所述标签向量进行分类，得到目标向量。

可选地，所述课程推荐装置还包括：

画像获取模块，用于获取每个样本用户的用户画像；

特征提取模块，用于提取每个所述用户画像的特征向量；

历史数据采集模块，用于获取每个所述样本用户的历史行为数据集合；

关联构建模块，用于确定每个所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别；

模型训练模块，用于针对每个所述样本用户，将所述样本用户的用户画像的特征向量作为输入，将所述样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练得到所述预先训练的课程推荐模型。

可选地，所述训练模块包括：

训练特征确定单元，用于将所述目标向量的每个维度作为一个训练特征；

决策树生成单元，用于将所述训练特征输入到梯度提升决策树模型中，通过梯度提升决策树模型对所述训练特征进行训练，得到n棵决策树，其中，n为正整数；

预测单元，用于将所每棵所述决策树的路径包含的训练特征作为自变量，基于二分类逻辑回归模型，进行课程类别的预测，得到每个课程类别对应的预测分值；

推荐类别确定单元，用于将预测分值超过预设阈值的课程类别，作为所述待推荐课程类别。

可选地，所述课程推荐装置还包括：

接收模块，用于接收所述待推荐用户针对所述目标推荐课程的学习评估数据；

画像更新模块，用于基于所述学习评估数据，对所述待推荐用户进行二次画像，得到更新画像；

推荐课程更新模块，用于基于所述更新画像，对所述目标推荐课程进行更新。

为了解决上述技术问题，本申请实施例还提供一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述课程推荐方法的步骤。

为了解决上述技术问题，本申请实施例还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述课程推荐方法的步骤。

本发明实施例提供的课程推荐方法、装置、计算机设备及存储介质，获取待推荐用户的基础数据，对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像，实现对每个待推荐用户进行精准定位，在后续通过用户画像对用户进行个性化推荐，有利于提高后续推荐的针对性和精准度，同时，根据用户画像进行特征提取，得到目标向量，将目标向量输入至预先训练的课程推荐模型中，并采用预先训练的课程推荐模型进行数据处理，得到待推荐课程类别，从待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将目标推荐课程推荐给待推荐用户，实现对每个待推荐用户，根据其自身画像特色，进行个性化推荐，有利于提高课程推荐的精准程度。

**附图说明**

为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1是本申请可以应用于其中的示例性系统架构图；

图2是本申请的课程推荐方法的一个实施例的流程图；

图3是根据本申请的课程推荐装置的一个实施例的结构示意图；

图4是根据本申请的计算机设备的一个实施例的结构示意图。

**具体实施方式**

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同；本文中在申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请；本申请的说明书和权利要求书及上述附图说明中的术语“包括”和“具有”以及它们的任何变形，意图在于覆盖不排他的包含。本申请的说明书和权利要求书或上述附图中的术语“第一”、“第二”等是用于区别不同对象，而不是用于描述特定顺序。

在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

请参阅图1，如图1所示，系统架构100可以包括终端设备101、102、103，网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。

终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、MP3播放器(Moving Picture E界面显示perts Group Audio Layer III，动态影像专家压缩标准音频层面3)、MP4(Moving PictureE界面显示perts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。

服务器105可以是提供各种服务的服务器，例如对终端设备101、102、103上显示的页面提供支持的后台服务器。

需要说明的是，本申请实施例所提供的课程推荐方法由服务器执行，相应地，课程推荐装置设置于服务器中。

应该理解，图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器，本申请实施例中的终端设备101、102、103具体可以对应的是实际生产中的应用系统。

请参阅图2，图2示出本发明实施例提供的一种课程推荐方法，以该方法应用在图1中的服务端为例进行说明，详述如下：

S201：获取待推荐用户的基础数据。

具体地，服务端在接收到待推荐用户的课程推荐请求时，获取课程推荐请求中包含的待推荐用户的用户标识，进而根据该用户标识获取待推荐用户的基础数据，以便后续根据获取到的基础数据，对待推荐用户进行精准画像，得到包含多个维度标签的待推荐用户的用户画像。

其中，基础数据是指用户的历史行为数据，具体可以是通过分布式大数据的方式，从各个系统爬取，进而存储到分布式数据中，历史行为数据具体包含的数据内容，可根据实际业务需求进行设定，例如，在业务推销领域，历史行为数据具体可以是各项指标进度、项目参与数据和成功率等。

S202：对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像。

对基础数据进行数据预处理和聚合的具体实现过程，可参考后续实施例的描述，为避免重复，此处不再赘述。

其中，用户画像(User Persona)的概念最早由交互设计之父Alan Cooper提出，是建立在一系列属性数据之上的目标用户模型。一般是产品设计、运营人员从用户群体中抽象出来的典型用户，本质是一个用以描述用户需求的工具。本实施例中，进行精准画像是根据对用户的历史行为数据进行分析，得出用户在各个维度的标签，形成多维度的用户画像。

其中，历史行为数据包括但不限于：历史搜索记录、历史点击记录、历史浏览记录等。历史搜索记录包括但不限于：用户信息、搜索时间和检索关键词。用户信息包括用户的基本信息，如姓名、性别、年龄等，搜索时间是指检测到检索操作的具体时间，检索关键词是指检索时间输入并进行查询的关键词。

需要说明的是，本实施例通过历史行为数据的分析，提取待推荐用户在处理业务过程中，对各个维度的技能掌握程度，从而生成各个维度的用户标签，得到用户画像，有利于后续通过精准画像实现课程精准推荐。

S203：根据用户画像进行特征提取，得到目标向量。

具体地，在步骤S202中，实现对用户进行精准画像后，得到具有多个维度标签的用户画像，本实施例中，对各个维度标签进行向量化和聚类，实现文字标签的数据化，后续可以通过将数字化的目标向量输入到预先训练的课程推荐模型中进行训练，得到推荐的课程信息，有利于提高课程推荐的准确率。

S204：将目标向量输入至预先训练的课程推荐模型中，并采用预先训练的课程推荐模型进行数据处理，得到待推荐课程类别。

具体地，在得到目标向量后，将目标向量输入到预先训练的课程推荐模型中，采用该模型进行训练，输出待推荐课程类别。

本实施例中，通过将多个样本用户的用户画像和样本用户对应的推荐课程类别作为初始数据，采用初始数据对神经网络模型或者机器学习模型中进行训练，得到预先训练的课程推荐模型。

S205：从待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将目标推荐课程推荐给待推荐用户。

具体地，从待推荐课程类别中，选取至少一门课程，作为待推荐课程，并生成课程推荐信息，发送给待推荐用户。

其中，课程推荐信息的生成，可以是通过向待推荐用户展示列表，根据接收到的用户选择信息来生成推荐信息，也可以结合待推荐用户的历史数据，对待推荐课程进行相关性排序，选取排序考前的预设个数课程生成课程推荐信息，此处不作具体限定。

本实施例中，获取待推荐用户的基础数据，对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像，实现对每个待推荐用户进行精准定位，在后续通过用户画像对用户进行个性化推荐，有利于提高后续推荐的针对性和精准度，同时，根据用户画像进行特征提取，得到目标向量，将目标向量输入至预先训练的课程推荐模型中，并采用预先训练的课程推荐模型进行数据处理，得到待推荐课程类别，从待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将目标推荐课程推荐给待推荐用户，实现对每个待推荐用户，根据其自身画像特色，进行个性化推荐，有利于提高课程推荐的精准程度。

在本实施例的一些可选的实现方式中，步骤S202中，对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像包括：

对基础数据进行数据预处理，得到标准数据；

对标准数据进行分词处理，并对分词处理结果进行聚类，得到待推荐用户的关联类别；

根据关联类别，生成目标标签和目标标签的重要性排序，得到用户画像。

具体地，在采集到用户的基础数据后，需要对数据进行脏数据清洗、数据完整性检查和规则化转换等处理，确保数据质量，得到标准数据，有利于提高得到用户画像的准确程度，同时，利用标准数据进行分词聚类，根据聚类结果确定用户的关联类别，进而根据对应的关联类别，确定目标标签和目标标签的排序，得到用户画像。

需要说明的是，每个待推荐用户的基础数据包括多个维度的多条数据，通过聚类，确定该待推荐用户主要数据分布领域，进而确定该待推荐用户对应的关联类别。进而根据落入该关联类别的数据，确定用户的目标标签。

其中，目标标签的重要性具体可依据目标标签与关联类别的相关性来进行确定，目标标签与关联类别越接近，则重要性越大。

本实施例中，通过对待推荐用户的基础数据进行数据分析，确定待推荐用户的关联类别，进而根据关联类别，生成用户画像，提高用户画像的精度。

在本实施例的一些可选的实现方式中，步骤S203中，根据用户画像进行特征提取，得到目标向量包括：

基于预设语料库，构建用户画像中每个标签对应的基础词向量；

针对每个基础词向量，计算该基础词向量与其他基础词向量之间的空间距离，并从空间距离中选取最小值作为该基础词向量的最小空间距离；

将最小空间距离中小于或等于预设空间距离阈值的基础词向量，作为标签向量；

基于K-Means聚合算法对标签向量进行分类，得到目标向量。

在人工智能中，语言表示主要指语言的形式化或数学的描述，以便在计算机中表示语言，并能让计算机程序自动处理。本发明实施例中所指的词向量就是用向量的形式来表示一个标签。

其中，预设语料库具体可以是根据实际业务需求，采集与业务相关联的语料得到。

具体地，通过使用词向量的方式对用户画像中每个标签进行训练，按照预设的语料库映射到一个向量中，将这些向量联系在一起，形成一个词向量空间，每个向量相当于是这个空间中的一个点得到符合预设要求的基础词向量，再通过对基础词向量进行筛选，得出关注度较高的标签向量，进一步对标签向量进行聚类，得到目标向量。

例如，在一具体实施方式中，预设的空间距离阈值为0.8，基础词向量包括H1(0.9,0.1,0)、H2(0.8,0.1,0.1)和H3(0,0.1,0.9)，通过计算各个基础词向量之间的欧式距离得到H1的最小空间距离为0.4243，到H2的最小空间距离为0.4243，到H3的最小空间距离为1.1314，H1和H2的最小空间距离小于预设的空间距离阈值0.8，因此，将H1和H2作为标签向量。

应理解，通过对不符合空间距离阈值要求的基础词向量进行过滤，避免了将关注度低的标签也放入用户标签，从而可以更准确的确定用户画像中业务的偏向和定位。

可以理解地，基础词向量的每一维的值代表一个具有一定的语义和语法上能够解释的特征，故可以将基础词向量的每一维称为一个标签特征。

其中，K-means算法是基于距离的聚类算法，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标，通过该聚类算法，对数量较多的标签向量进行精炼，有利于提高后续模型训练推荐的效率。

其中，预设空间距离阈值可根据实际需求进行设置，此处不做限定。

在本实施例中，构建用户画像的每个标签对应的基础词向量，进而进行聚类，确定用户画像的类别，得到目标向量，有利于后续通过目标向量，参与模型训练，快速得到推荐课程类别。

在本实施例的一些可选的实现方式中，步骤S204之前，该课程推荐方法还包括：

获取每个样本用户的用户画像；

提取每个用户画像的特征向量；

获取每个样本用户的历史行为数据集合；

确定每个样本用户的历史行为数据集合中的历史行为数据所指示的关联类别；

针对每个样本用户，将样本用户的用户画像的特征向量作为输入，将样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练得到预先训练的课程推荐模型。

其中，历史行为数据集合是指通过对样本用户的历史行为数据进行收集和分析，得到的样本用户的历史行为相关的数据集合，根据该数据集合，确定样本用户对应的关联类别，进而将用户画像对应的特征向量作为输入，将关联类别作为输出，对初始的课程推荐模型进行有监督学习训练，得到预先训练的课程推荐模型。

其中，历史行为数据包括但不限于：历史搜索记录、历史点击记录、历史浏览记录等。历史搜索记录包括但不限于：用户信息、搜索时间和检索关键词。用户信息包括用户的基本信息，如姓名、性别、年龄等，搜索时间是指检测到检索操作的具体时间，检索关键词是指检索时间输入并进行查询的关键词。

在本实施例中，基于上述步骤所提取的多个样本用户的用户画像的特征向量和所确定的多个样本用户的历史行为数据集合中的历史行为数据所指示的关联类别，服务端可以将多个样本用户的用户画像的特征向量作为输入，将多个样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练出能够表征特征向量与关联类别之间对应关系的课程推荐模型。

其中，关联类别是指历史行为数据对应的课程推荐类别，例如，捕获到某一历史行为为操作PPT表格，对其数据进行评估，评估结果为该技能熟练度为中级，进而确定推荐的关联类别为中级或高级PPT课程。

在本实施例中，课程推荐模型可以是各种类型的人工神经网络或者对多种类型的人工神经网络进行组合所得到的模型。它从信息处理角度对人脑神经元网络进行抽象，建立某种简单模型，按不同的连接方式组成不同的网络。通常由大量的节点(或称神经元)之间相互联接构成，每个节点代表一种特定的输出函数，称为激励函数。每两个节点间的连接都代表一个对于通过该连接信号的加权值，称之为权重，网络的输出则依网络的连接方式、权重值和激励函数的不同而不同。这里，服务端训练的可以是初始化课程推荐模型，初始化课程推荐模型可以是未经训练的课程推荐模型或未训练完成的课程推荐模型，初始化课程推荐模型可以设置有初始参数，参数在课程推荐模型的训练过程中可以被不断地调整，直至训练出能够表征特征向量与推荐课程类别之间对应关系的课程推荐模型为止。

在本实施例中，服务端可以将样本用户的历史行为数据集合中的历史行为数据所指示的关联类别全部作为推荐课程类别，并将推荐课程类别作为输出；也可以从样本用户的历史行为数据集合中的历史行为数据所指示的关联类别中选取出部分类别作为推荐课程类别，并将推荐课程类别作为输出；还可以从样本用户的历史行为数据集合中的历史行为数据所指示的关联类别中选取出部分类别作为推荐课程类别，选取出另一部分类别作为非推荐课程类别，并将推荐课程类别和非推荐课程类别作为输出，其中，推荐课程类别可以是样本用户感兴趣的类别(与用户的历史行为数据关联密切的类别)，非推荐课程类别可以是样本用户不感兴趣的类别。

本实施例中，通过预先对课程推荐模型进行训练，得到预先训练的课程推荐模型，在后续使用时直接采用预先训练的课程推荐模型对待推荐用户的目标向量进行训练，有利于提高模型训练的效率，同时，采用相同场景的多个真实数据作为训练数据，也有利于提高该预先训练的课程推荐模型的推荐准确率。

在本实施例的一些可选的实现方式中，步骤S205中可选地，预先训练的课程推荐模型为梯度提升决策树模型，采用预先训练的课程推荐模型进行数据处理，得到待推荐课程类别包括：

将目标向量的每个维度作为一个训练特征；

将训练特征输入到梯度提升决策树模型中，通过梯度提升决策树模型对训练特征进行训练，得到n棵决策树，其中，n为正整数；

将所每棵决策树的路径包含的训练特征作为自变量，基于二分类逻辑回归模型，进行课程类别的预测，得到每个课程类别对应的预测分值；

将预测分值超过预设阈值的课程类别，作为待推荐课程类别。

其中，梯度提升决策树(Gradient Boosting Decision Tree，GBDT)算法是一种迭代的决策树算法，该算法由多棵决策树组成，所有树的结论累加起来作为最终预测结果。

其中，梯度提升决策树中的决策树属于回归树，在这些树的每个节点都会得到该节点对应的分类特征的预测值，对于未确定具体数值的分类特征，使用该分类特征的平均值作为该分类特征的预测值。

在生成决策树模型后，针对每个决策树，将不同路径包含的特征的特征值进行特征组合，得到组合特征，并将不同树的相同组合特征的值进行累加，将最终的累加值作为该组合特征的特征值，并将该特征值作为二分类逻辑回归(Logistic Regression，LR)模型中的一个自变量，来计算概率值，并根据概率值确定组合特征对应的目标向量的预测分值，进而根据预测分值，确定待推荐课程类别，有利于提高待推荐课程类别推荐的准确性。

在本实施例的一些可选的实现方式中，步骤S205之后中，课程推荐方法还包括：

接收待推荐用户针对目标推荐课程的学习评估数据；

基于学习评估数据，对待推荐用户进行二次画像，得到更新画像；

基于更新画像，对目标推荐课程进行更新。

具体地，在待推荐用户对淼推荐课程进行一段时间的学习后，通过接收待推荐用户学习后的评估数据(测试数据)，对待推荐用户进行二次画像，具体过程可参考上述实施例，实现用户画像的及时更新，进而根据实际需求，判断是否需要进行目标推荐课程的更新。

本实施例中，通过对目标推荐课程学习后进行二次画像评估，判断是否需要根据学习结果进行目标推荐课程的更新，进一步提高课程推荐的精准程度。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

图3示出与上述实施例课程推荐方法一一对应的课程推荐装置的原理框图。如图3所示，该课程推荐装置包括采集模块31、画像模块32、提取模块33、训练模块34和推荐模块35。各功能模块详细说明如下：

采集模块31，用于获取待推荐用户的基础数据；

画像模块32，用于对基础数据进行数据预处理和聚合，根据得到的聚合结果，对待推荐用户进行画像处理，得到用户画像；

提取模块33，用于根据用户画像进行特征提取，得到目标向量；

训练模块34，用于将目标向量输入至预先训练的课程推荐模型中，并采用预先训练的课程推荐模型进行数据处理，得到待推荐课程类别；

推荐模块35，用于从待推荐课程类别中，选取至少一个课程推荐信息，作为目标推荐课程，并将目标推荐课程推荐给待推荐用户。

可选地，画像模块32包括：

预处理单元，用于对基础数据进行数据预处理，得到标准数据；

聚类单元，用于对标准数据进行分词处理，并对分词处理结果进行聚类，得到待推荐用户的关联类别；

画像生成单元，用于根据关联类别，生成目标标签和目标标签的重要性排序，得到用户画像。

可选地，提取模块33包括：

基础词向量构建单元，用于基于预设语料库，构建用户画像中每个标签对应的基础词向量；

空间距离计算单元，用于针对每个基础词向量，计算该基础词向量与其他基础词向量之间的空间距离，并从空间距离中选取最小值作为该基础词向量的最小空间距离；

词向量筛选单元，用于将最小空间距离中小于或等于预设空间距离阈值的基础词向量，作为标签向量；

分类单元，用于基于K-Means聚合算法对标签向量进行分类，得到目标向量。

可选地，课程推荐装置还包括：

画像获取模块，用于获取每个样本用户的用户画像；

特征提取模块，用于提取每个用户画像的特征向量；

历史数据采集模块，用于获取每个样本用户的历史行为数据集合；

关联构建模块，用于确定每个样本用户的历史行为数据集合中的历史行为数据所指示的关联类别；

模型训练模块，用于针对每个样本用户，将样本用户的用户画像的特征向量作为输入，将样本用户的历史行为数据集合中的历史行为数据所指示的关联类别作为输出，训练得到预先训练的课程推荐模型。

可选地，训练模块包括：

训练特征确定单元，用于将目标向量的每个维度作为一个训练特征；

决策树生成单元，用于将训练特征输入到梯度提升决策树模型中，通过梯度提升决策树模型对训练特征进行训练，得到n棵决策树，其中，n为正整数；

预测单元，用于将所每棵决策树的路径包含的训练特征作为自变量，基于二分类逻辑回归模型，进行课程类别的预测，得到每个课程类别对应的预测分值；

推荐类别确定单元，用于将预测分值超过预设阈值的课程类别，作为待推荐课程类别。

可选地，课程推荐装置还包括：

接收模块，用于接收待推荐用户针对目标推荐课程的学习评估数据；

画像更新模块，用于基于学习评估数据，对待推荐用户进行二次画像，得到更新画像；

推荐课程更新模块，用于基于更新画像，对目标推荐课程进行更新。

关于课程推荐装置的具体限定可以参见上文中对于课程推荐方法的限定，在此不再赘述。上述课程推荐装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

为解决上述技术问题，本申请实施例还提供计算机设备。具体请参阅图4，图4为本实施例计算机设备基本结构框图。

所述计算机设备4包括通过系统总线相互通信连接存储器41、处理器42、网络接口43。需要指出的是，图中仅示出了具有组件连接存储器41、处理器42、网络接口43的计算机设备4，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。其中，本技术领域技术人员可以理解，这里的计算机设备是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit，ASIC)、可编程门阵列(Field－Programmable Gate Array，FPGA)、数字处理器(Digital Signal Processor，DSP)、嵌入式设备等。

所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

所述存储器41至少包括一种类型的可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如，SD或D界面显示存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中，所述存储器41可以是所述计算机设备4的内部存储单元，例如该计算机设备4的硬盘或内存。在另一些实施例中，所述存储器41也可以是所述计算机设备4的外部存储设备，例如该计算机设备4上配备的插接式硬盘，智能存储卡(Smart Media Card,SMC)，安全数字(Secure Digital,SD)卡，闪存卡(Flash Card)等。当然，所述存储器41还可以既包括所述计算机设备4的内部存储单元也包括其外部存储设备。本实施例中，所述存储器41通常用于存储安装于所述计算机设备4的操作系统和各类应用软件，例如电子文件的控制的程序代码等。此外，所述存储器41还可以用于暂时地存储已经输出或者将要输出的各类数据。

所述处理器42在一些实施例中可以是中央处理器(Central Processing Unit，CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器42通常用于控制所述计算机设备4的总体操作。本实施例中，所述处理器42用于运行所述存储器41中存储的程序代码或者处理数据，例如运行电子文件的控制的程序代码。

所述网络接口43可包括无线网络接口或有线网络接口，该网络接口43通常用于在所述计算机设备4与其他电子设备之间建立通信连接。

本申请还提供了另一种实施方式，即提供一种计算机可读存储介质，所述计算机可读存储介质存储有界面显示程序，所述界面显示程序可被至少一个处理器执行，以使所述至少一个处理器执行如上述的课程推荐方法的步骤。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，空调器，或者网络设备等)执行本申请各个实施例所述的方法。

显然，以上所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例，附图中给出了本申请的较佳实施例，但并不限制本申请的专利范围。本申请可以以许多不同的形式来实现，相反地，提供这些实施例的目的是使对本申请的公开内容的理解更加透彻全面。尽管参照前述实施例对本申请进行了详细的说明，对于本领域的技术人员来而言，其依然可以对前述各具体实施方式所记载的技术方案进行修改，或者对其中部分技术特征进行等效替换。凡是利用本申请说明书及附图内容所做的等效结构，直接或间接运用在其他相关的技术领域，均同理在本申请专利保护范围之内。

|  |
| --- |
| **说 明 书 附 图** |

