(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118570030 A
(43)申请公布日 2024.08.30

(21)申请号 202411000195.X

(22)申请日 2024.07.24

(71)申请人 安徽教育出版社
地址 230061 安徽省合肥市经济技术开发区繁华大道398号
申请人 安徽奇初教育科技有限公司

(72)发明人 李冰冰 宋潇婧 章红红 李晨希 郑娟 庞飞天 程桃莉

(74)专利代理机构 上海复暨知识产权代理事务所(普通合伙) 31449
专利代理师 刘东亮

(51)Int.Cl.
G06Q 50/20(2012.01)
G06F 18/25(2023.01)
G06N 5/022(2023.01)
G06N 3/0455(2023.01)
G06N 3/094(2023.01)
G06F 18/23213(2023.01)
G06F 40/284(2020.01)
G06F 40/30(2020.01)

权利要求书3页 说明书13页 附图3页

(54)发明名称

基于多源数据融合的作业生成方法及系统

(57)摘要

本发明提供了基于多源数据融合的作业生成方法与系统,涉及教育技术领域,该方法通过获取多源异构数,处理后得到全面覆盖学生学习情况的多源数据,提取语义信息,分解得到语义表征向量后融合后得到多源学习数据,构建学生画像,识别出每个学生的知识薄弱点和学习兴趣点,训练作业指导模型,逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集,从资源子数据集中推送依据所述关键词数据筛选的资源数据,生成符合每个学生的个性化作业。本发明通过融合多源异构数据,精准识别学生的知识薄弱点和学习兴趣,生成个性化的学习作业,以提升学习效果。

CN 118570030 A

1.一种基于多源数据融合的作业生成方法,其特征在于,该方法包括以下步骤:

获取用于数据融合的多源异构数据,对多源异构数据进行标准化处理后,得到全面覆盖学生学习情况的多源数据;

将得到的多源数据转化为文本数据,根据语义知识库提取多源数据中的语义信息,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,并对语义表征向量进行融合,得到融合后的多源学习数据;

根据融合后的多源学习数据构建学生画像,识别出每个学生的知识薄弱点和学习兴趣点;

以语义表征向量和包括知识薄弱点和学习兴趣点的学生画像为输入,以个性化作业推荐为标签,训练作业指导模型;

逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集;

基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中推送依据所述关键词数据筛选的资源数据,生成符合每个学生的个性化作业。

2.如权利要求1所述的基于多源数据融合的作业生成方法,其特征在于,获取的多源异构数据包括学生的课堂表现数据、考试成绩数据、作业完成情况数据、在线学习数据以及学习行为数据。

3.如权利要求2所述的基于多源数据融合的作业生成方法,其特征在于,对多源异构数据进行标准化处理时,还包括对多源异构数据进行插值和偏差订正处理,其中,采用反距离权重插值将多源异构数据插值到指定分辨率的网格上,使用历史数据对当前数据进行偏差订正,插值时,插值点的估计值计算公式为:$V(x)=\frac{\sum_{i-1}^{N}\frac{V_i}{d_i^p}}{\sum_{i-1}^{N}\frac{1}{d_i^p}}$式中,$V(x)$为插值点$x$处的估计值,$N$为参与插值计算的已知数据点的总数,$i$是从1到$N$的索引,用于标记每一个已知数据点,$V_i$为第$i$个已知点的数值,$d_i$为插值点与第$i$个已知点之间的距离,$p$为权重指数。

4.如权利要求3所述的基于多源数据融合的作业生成方法,其特征在于,插值点x的估计值$V(x)$是已知点值V_i的加权平均值,权重是已知点与插值点之间距离的倒数的p次幂,其中,$V(x)$的计算公式中,分子$\sum_{i-1}^{N}\frac{V_i}{d_i^p}$为对所有已知点值$V_i$按照其权重$\frac{1}{d_i^p}$进行加权求和,分母$\sum_{i-1}^{N}\frac{1}{d_i^p}$为计算所有已知点权重的总和,估计值$V(x)$为分子除以分母,得到插值点$x$处的估计值$V(x)$。

5.如权利要求4所述的基于多源数据融合的作业生成方法,其特征在于,根据语义知识库提取多源数据中的语义信息时,基于现有的WordNet、DBpedia和ConceptNet知识库,选择

和构建覆盖多源异构数据领域的语义知识库;应用Stanford NER在标准化处理后的多源数据上识别出命名实体,并将识别出的实体与语义知识库中的条目进行匹配,在识别出的实体上进行词义消歧,于语义知识库提取实体和概念之间的语义关系,构建和生成包含实体和关系的语义图谱。

6.如权利要求5所述的基于多源数据融合的作业生成方法,其特征在于,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,包括:

将提取语义信息后的文本数据输入预训练的词向量模型,生成对应的词向量;

将词向量映射到一个共享的语义空间,使用对抗训练对齐不同源的词向量;

将对齐后的词向量组合构建形成完整的词向量矩阵;

对词向量矩阵进行中心化处理,减去每一列的均值,应用奇异值分解将词向量矩阵分解为三个矩阵U、Σ、V^T;选取前k个奇异值及其对应的奇异向量,得到低维语义表征向量;

使用分解后的U矩阵中的前k列,作为词的低维语义表征向量。

7.如权利要求6所述的基于多源数据融合的作业生成方法,其特征在于,所述词向量模型为预训练的Word2Vec模型,通过使用Word2Vec在语料库上进行训练得到,对抗训练时,还包括设计的一个对抗训练框架,该对抗训练框架包括生成器和判别器,生成器负责将词向量转换到共享语义空间,判别器则评估词向量是否来自同一数据源。

8.如权利要求7所述的基于多源数据融合的作业生成方法,其特征在于,构建学生画像时,包括:

构建包含知识水平、学习习惯、兴趣爱好多维特征的学生画像,将融合后的多源学习数据分解成多维的特征子集,建立知识点映射矩阵;

进行知识点掌握度计算,计算每个学生对各知识点的掌握度;

设定一个掌握度阈值,识别出掌握度低于阈值的知识点;

使用K-means对学习兴趣特征子集进行聚类,识别出不同的兴趣点,对聚类结果进行标注,识别出每个兴趣点的含义;

将上述计算结果整合,生成每个学生的详细画像报告。

9.如权利要求8所述的基于多源数据融合的作业生成方法,其特征在于,训练作业指导模型时,将生成的语义表征向量与对应的语义标签配对,形成训练数据集模型进行训练;其中,作业语义表征向量作为输入,知识薄弱点和学习兴趣点对应的语义信息作为标签;选择Transformer模型并使用生成的训练数据集进行模型训练,得到作业指导模型,利用训练好的模型生成个性化作业指导。

10.一种基于多源数据融合的作业生成系统,其特征在于,用于执行权利要求1-9中任意一项所述基于多源数据融合的作业生成方法,所述基于多源数据融合的作业生成系统包括:

多源数据获取模块:用于获取多源异构数据,并对获取的多源异构数据进行标准化处理,生成全面覆盖学生学习情况的多源数据;

语义信息提取模块:用于将标准化处理后的多源数据转化为文本数据,根据语义知识库,提取多源数据中的语义信息,并投影到共享语义空间,生成词向量矩阵;

词向量分解及融合模块:用于对词向量矩阵进行分解,得到语义表征向量,并对这些向量进行融合,生成融合后的多源学习数据;

学生画像构建模块:基于融合后的多源学习数据,构建学生画像,识别每个学生的知识薄弱点和学习兴趣点;

作业指导模型训练模块:用于以语义表征向量为输入,以对应的语义信息为标签,结合学生画像及其知识薄弱点和学习兴趣点,训练作业指导模型;

在线内容识别模块:用于逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据;

架构树生成模块:用于根据识别出的关键词数据,生成对应的资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集;

作业生成模块:用于基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中筛选资源数据,生成符合每个学生的个性化作业。

基于多源数据融合的作业生成方法及系统

技术领域

[0001] 本发明涉及教育技术领域,具体涉及基于多源数据融合的作业生成方法与系统。

背景技术

[0002] 随着信息技术的不断发展和教育资源的日益丰富,教育领域正在经历一场深刻的变革。传统的教育模式主要依赖于教师的经验和教材内容,作业内容通常是统一设计,缺乏个性化和针对性,这在一定程度上制约了学生的学习效果和教学质量的提升。在这种背景下,利用大数据、人工智能和数据融合技术生成个性化的作业内容,逐渐成为教育技术领域的研究热点。

[0003] 其中,传统作业生成方式存在缺乏个性化、效率低下以及反馈滞后的局限性。原因在于,传统的作业主要基于教材和教师的经验布置,难以针对每个学生的学习水平和需求进行调整,造成部分学生作业难度不适应其学习状态,影响学习效果。其次,教师需要花费大量时间和精力设计作业,检查和反馈作业结果,这在大班教学中显得尤为困难。而且,学生在完成作业后,需要等待教师批改和反馈,这一过程很可能导致错过最佳的学习纠错时间。

[0004] 而随着教育信息化的发展,在线学习平台和学习管理系统也在不断普及。而随着大量在线学习平台涌现,这些学习平台可以记录学生详细的学习行为数据,如观看视频的时间、做题情况、课程完成度等,为个性化教育提供了数据支持。同时,学习管理系统也可以方便地管理课程内容、学生记录和学习活动数据。这些系统为收集和分析学生的学习行为提供了基础设施。

[0005] 但是目前的在线学习平台和学习管理系统虽然能够在线记录学生学习情况和课程管理,但是覆盖学生学习进度及学习情况的数据来源广泛,包括但不局限于在线学习平台、课堂表现、考试成绩、教师评价等。如何将这些异构数据有效融合,是实现个性化教育的关键。尤其是如何根据每个学生的特点和需求,提供适合其学习路径的作业内容,是有效提升学生学习效率和效果的关键。

发明内容

[0006] 有鉴于此,针对上述问题,本发明提出了一种基于多源数据融合的作业生成方法与系统,旨在通过融合多源异构数据,精准识别学生的知识薄弱点和学习兴趣,生成个性化的学习作业,以提升学习效果。

[0007] 本发明采用以下技术方案实现:

第一方面,本发明提供了一种基于多源数据融合的作业生成方法,该方法包括以下步骤:

获取用于数据融合的多源异构数据,对多源异构数据进行标准化处理后,得到全面覆盖学生学习情况的多源数据;

将得到的多源数据转化为文本数据,根据语义知识库提取多源数据中的语义信

息,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,并对语义表征向量进行融合,得到融合后的多源学习数据;

根据融合后的多源学习数据构建学生画像,识别出每个学生的知识薄弱点和学习兴趣点;

以语义表征向量和包括知识薄弱点和学习兴趣点的学生画像为输入,以个性化作业推荐为标签,训练作业指导模型;

逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集;

基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中推送依据所述关键词数据筛选的资源数据,生成符合每个学生的个性化作业。

[0008] 通过上述步骤,本发明能够有效地利用多源数据融合生成符合学生个性化需求的作业,帮助学生更好地掌握知识点并提升学习兴趣和效果。

[0009] 作为本发明的进一步方案,获取的多源异构数据包括但不限于:

学生的课堂表现数据(如出勤率、课堂参与度等)。

[0010] 学生的考试成绩数据(如期中考试、期末考试等)。

[0011] 学生的作业完成情况数据(如作业完成时间、正确率等)。

[0012] 学生的在线学习数据(如在线课程观看记录、在线测试结果等)。

[0013] 学生的学习行为数据(如学习时间分布、学习路径等)。

[0014] 作为本发明的进一步方案,对多源异构数据进行标准化处理时,还包括对多源异构数据进行插值和偏差订正处理,其中,采用反距离权重插值将多源异构数据插值到指定分辨率的网格上,使用历史数据对当前数据进行偏差订正,插值时,插值点的估计值计算公式为:$V(x)=\frac{\sum_{i-1}^{N}\frac{V_i}{d_i^p}}{\sum_{i-1}^{N}\frac{1}{d_i^p}}$式中,$V(x)$为插值点$x$处的估计值,$N$为参与插值计算的已知数据点的总数,$i$是从1到$N$的索引,用于标记每一个已知数据点,$V_i$为第$i$个已知点的数值,$d_i$为插值点与第$i$个已知点之间的距离,$p$为权重指数。

[0015] 其中,插值点x的估计值$V(x)$是已知点值V_i的加权平均值,权重是已知点与插值点之间距离的倒数的p次幂,其中,$V(x)$的计算公式中,分子$\sum_{i-1}^{N}\frac{V_i}{d_i^p}$为对所有已知点值$V_i$按照其权重$\frac{1}{d_i^p}$进行加权求和,分母$\sum_{i-1}^{N}\frac{1}{d_i^p}$为计算所有已知点权重的总和,估计值$V(x)$为分子除以分母,得到插值点$x$处的估计值$V(x)$。

[0016] 作为本发明的进一步方案,根据语义知识库提取多源数据中的语义信息时,基于现有的WordNet、DBpedia和ConceptNet知识库,选择和构建覆盖多源异构数据领域的语义知识库;应用Stanford NER在标准化处理后的多源数据上识别出命名实体,并将识别出的

实体与语义知识库中的条目进行匹配,在识别出的实体上进行词义消歧,于语义知识库提取实体和概念之间的语义关系,构建和生成包含实体和关系的语义图谱。通过以上详细步骤,可以系统化地从多源数据中提取语义信息,生成准确且结构清晰的语义图谱。每个步骤都有明确的执行方式,并且通过权利要求书保证了方法的创新性和合法性。

[0017] 作为本发明的进一步方案,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,包括:

将提取语义信息后的文本数据输入预训练的词向量模型,生成对应的词向量;

将词向量映射到一个共享的语义空间,使用对抗训练对齐不同源的词向量;

将对齐后的词向量组合构建形成完整的词向量矩阵;

对词向量矩阵进行中心化处理,减去每一列的均值,应用奇异值分解将词向量矩阵分解为三个矩阵U、Σ、V^T,其中:$A = U\Sigma V^T$ 式中,A是一个m×n的词向量矩阵;U是一个m×n的正交矩阵,U列向量是A的左奇异向量;Σ是一个 m×n的对角矩阵,Σ对角线上的元素是A的奇异值;V^T是一个 m×n的正交矩阵,V^T行向量是A的右奇异向量的转置;

选取前k个奇异值及其对应的奇异向量,得到低维语义表征向量;选取前k个奇异值后,公式表示为:$A_k \approx U_k\Sigma_k V_k^T$ 式中,U_k是U的前k列,大小m×k;Σ_k是Σ的前k个奇异值,对应一个k×k的对角矩阵;V_k^T是V^T的前k行,大小为k×n;使用分解后的U矩阵中的前k列,作为词的低维语义表征向量。

[0018] 其中,所述词向量模型为预训练的Word2Vec模型,通过使用Word2Vec在语料库上进行训练得到,对抗训练时,还包括设计的一个对抗训练框架,该对抗训练框架包括生成器和判别器,生成器负责将词向量转换到共享语义空间,判别器则评估词向量是否来自同一数据源,通过反复训练生成器和判别器,使不同源的词向量在共享语义空间中具有一致性。

[0019] 通过上述步骤,可以将多源数据中的词向量对齐到共享语义空间,并通过奇异值分解方法生成低维语义表征向量,这一过程确保了不同来源的数据能够在统一的语义空间中进行比较和分析,提供了高效且准确的语义表征。

[0020] 作为本发明的进一步方案,构建学生画像时,构建包含知识水平、学习习惯、兴趣爱好多维特征的学生画像,将融合后的多源学习数据$\mathbf{Z}$分解成多维的特征子集,建立知识点映射矩阵:$M = \begin{bmatrix} m_{11}m_{12}\cdots m_{1j} \\ m_{21}m_{22}\cdots m_{2j} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ m_{i1}m_{i2}\cdots m_{ij} \end{bmatrix}$ 式中,m_{ij}表示第i个学习活动涉及第j个知识点的程度;进行知识点掌握度计算,计算每个学生对各知识点的掌握度:$K = Z_k \times M$ 其中,K是学生对各知识点的掌握度矩阵,其中,Z_k为分解成的特征子集;

薄弱点识别,设定一个掌握度阈值θ,识别出掌握度低于阈值的知识点:$W = \{k_i | k_i < \theta\}$式中,W是学生的知识薄弱点集合;使用K-means对学习兴趣特征子集进行聚类,识别出不同的兴趣点,对聚类结果进行标注,识别出每个兴趣点的含义;

将上述计算结果整合,生成每个学生的详细画像报告。

[0021] 作为本发明的进一步方案,训练作业指导模型时,将生成的语义表征向量与对应的语义标签配对,形成训练数据集模型进行训练;其中,作业语义表征向量作为输入,知识薄弱点和学习兴趣点对应的语义信息作为标签;选择Transformer模型并使用生成的训练数据集进行模型训练,得到作业指导模型,利用训练好的模型生成个性化作业指导。

[0022] 作为本发明的进一步方案,生成知识点架构树时,根据章节内容和提取的关键词,初步建立知识点架构树的层级结构,通过语义分析和关联规则挖掘,确定各知识点之间的关系,去除冗余节点,优化架构树结构,将知识点与资料数据库中的资源进行匹配,根据得到的知识点架构树筛选出与其对应的资源子数据集。

[0023] 第二方面,本发明还包括一种基于多源数据融合的作业生成系统,该系统包括:

多源数据获取模块:用于获取多源异构数据,并对获取的多源异构数据进行标准化处理,生成全面覆盖学生学习情况的多源数据;

语义信息提取模块:用于将标准化处理后的多源数据转化为文本数据,根据语义知识库,提取多源数据中的语义信息,并投影到共享语义空间,生成词向量矩阵;

词向量分解及融合模块:用于对词向量矩阵进行分解,得到语义表征向量,并对这些向量进行融合,生成融合后的多源学习数据;

学生画像构建模块:基于融合后的多源学习数据,构建学生画像,识别每个学生的知识薄弱点和学习兴趣点;

作业指导模型训练模块:用于以语义表征向量为输入,以对应的语义信息为标签,结合学生画像及其知识薄弱点和学习兴趣点,训练作业指导模型;

在线内容识别模块:用于逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据;

架构树生成模块:用于根据识别出的关键词数据,生成对应的资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集;

作业生成模块:用于基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中筛选资源数据,生成符合每个学生的个性化作业。

[0024] 作为本发明的进一步方案,该作业生成系统还包括数据推送模块,用于将生成的个性化作业推送至学生端,确保每个学生都能收到符合其学习需求和兴趣的作业。

[0025] 通过上述模块的协同工作,本发明的作业生成系统能够高效、准确地生成个性化的学习作业,帮助学生更好地掌握知识,提升学习效果。

[0026] 本发明还包括一种计算机设备,包括:至少一个处理器,以及与所述至少一个处理器通信连接的存储器,其中,所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器执行所述的基于多源数据融合的作业生成方法。

[0027] 本发明还包括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指令,所述计算机指令用于使所述计算机执行所述的基于多源数据融合的作业生成方法。

[0028] 与现有技术相比,本发明提供的基于多源数据融合的作业生成方法与系统,具有以下有益效果:

1.通过获取和融合多源异构数据,如学术成绩、课堂表现、作业成绩、在线学习记录等,全面覆盖学生的学习情况,有助于准确评估学生的学习状态,消除不同数据源之间的

差异后,得到全面覆盖学生学习情况的多源数据。

[0029] 2.实现了语义信息提取和多源数据融合。利用语义知识库提取多源数据中的语义信息,并投影到共享语义空间,生成词向量矩阵,通过分解和融合生成语义表征向量,确保数据中的重要信息能够被充分利用。

[0030] 3.能够构建个性化学生画像和作业指导模型。基于融合后的多源学习数据,构建详细的学生画像,识别每个学生的知识薄弱点和学习兴趣点,提供了个性化学习的基础;以语义表征向量为输入,结合学生画像及其知识薄弱点和学习兴趣点,训练精确的作业指导模型,确保生成的作业能够针对学生的具体需求和兴趣。

[0031] 4.实现了在线内容的智能识别与资源数据筛选,针对性地生成作业。通过逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应的知识点架构树,确保作业内容与学生实际学习内容的高度相关性。基于生成的知识点架构树和学生画像,从资源子数据集中筛选出最适合的资源数据,确保学生获得的作业内容是最契合其学习需求的。通过作业指导模型和学生画像,生成符合每个学生的个性化作业,帮助学生更有针对性地进行学习,提升学习效果。

[0032] 综上所述,本发明的基于多源数据融合的作业生成方法与系统通过全面的数据获取、智能的语义分析、精准的作业指导和高效的个性化作业生成,显著提升了作业生成的准确性和个性化程度,促进了学生的学习效果和教育质量的提升。

[0033] 本发明的这些方面或其他方面在以下实施例的描述中会更加简明易懂。应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本发明。

附图说明

[0034] 为了更清楚地说明本发明实施例或相关技术中的技术方案,下面将对示例性实施例或相关技术描述中所需要使用的附图作一简单地介绍,附图用来提供对本发明的进一步理解,并且构成说明书的一部分,与本发明的实施例一起用于解释本发明,并不构成对本发明的限制。在附图中:

图1为本发明实施例的基于多源数据融合的作业生成方法的流程图。

[0035] 图2为本发明实施例的基于多源数据融合的作业生成方法中分解得到语义表征向量的流程图。

[0036] 图3为本发明实施例的基于多源数据融合的作业生成方法中构建学生画像的流程图。

具体实施方式

[0037] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0038] 在本发明的说明书和权利要求书及上述附图中的描述的一些流程中,包含了按照特定顺序出现的多个操作,但是应该清楚了解,这些操作可以不按照其在本文中出现的顺序来执行或并行执行,操作的序号如101、102等,仅仅是用于区分开各个不同的操作,序号本身不代表任何的执行顺序。另外,这些流程可以包括更多或更少的操作,并且这些操作可

以按顺序执行或并行执行。需要说明的是,本文中的“第一”、“第二”等描述,是用于区分不同的消息、设备、模块等,不代表先后顺序,也不限定“第一”和“第二”是不同的类型。

[0039] 下面将结合本发明示例性实施例中的附图,对本发明示例性实施例中的技术方案进行清楚、完整地描述,显然,所描述的示例性实施例仅仅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。

[0040] 为了根据每个学生的特点和需求,提供适合其学习路径的作业内容,本发明提供的一种基于多源数据融合的作业生成方法与系统,通过融合多源异构数据,精准识别学生的知识薄弱点和学习兴趣,生成个性化的学习作业,以提升学习效果。

[0041] 下面结合具体实施例对本发明的技术方案作进一步的说明:

参阅图1所示,图1为本发明提供的一种基于多源数据融合的作业生成方法的流程图。本发明一个实施例中提供的一种基于多源数据融合的作业生成方法,包括以下步骤:

步骤S10、获取用于数据融合的多源异构数据,对多源异构数据进行标准化处理后,得到全面覆盖学生学习情况的多源数据。

[0042] 该步骤中,获取的多源异构数据包括但不限于:

学生的课堂表现数据(如出勤率、课堂参与度等)。

[0043] 学生的考试成绩数据(如期中考试、期末考试等)。

[0044] 学生的作业完成情况数据(如作业完成时间、正确率等)。

[0045] 学生的在线学习数据(如在线课程观看记录、在线测试结果等)。

[0046] 学生的学习行为数据(如学习时间分布、学习路径等)。

[0047] 示例性的,在通过本实施例基于多源数据融合的方法生成个性化作业时,在选定的某一学期中,获取了某个学生(小明)的以下多源异构数据:

1.课堂表现数据:

出勤率:95%;

课堂参与度:高;

2.考试成绩数据:

期中考试成绩:85分;

期末考试成绩:88分;

3.作业完成情况数据:

平均作业完成时间:40分钟;

平均正确率:90%;

4.在线学习数据:

在线课程观看记录:完成率80%;

在线测试结果:85%;

5.学习行为数据:

学习时间分布:每天晚上7点到9点;

学习路径:先复习课堂笔记,再完成作业,最后进行在线测试。

[0048] 在本实施例中,对多源异构数据进行标准化处理时,还包括对多源异构数据进行插值和偏差订正处理,其中,采用反距离权重插值将多源异构数据插值到指定分辨率的网

格上,使用历史数据对当前数据进行偏差订正,插值时,插值点的估计值计算公式为:

$V(x)=\frac{\sum_{i-1}^{N}\frac{V_i}{d_i^p}}{\sum_{i-1}^{N}\frac{1}{d_i^p}}$ 式中,$V(x)$为插值点x处的估计值,N为参与插值计算的已知数据点的总数,i是从1到N的索引,用于标记每一个已知数据点,V_i为第i个已知点的数值,d_i为插值点与第i个已知点之间的距离,p为权重指数。其中,插值点x的估计值$V(x)$是已知点值V_i的加权平均值,权重是已知点与插值点之间距离的倒数的p次幂,其中,$V(x)$的计算公式中,分子$\sum_{i-1}^{N}\frac{V_i}{d_i^p}$为对所有已知点值$V_i$按照其权重$\frac{1}{d_i^p}$进行加权求和,分母$\sum_{i-1}^{N}\frac{1}{d_i^p}$为计算所有已知点权重的总和,估计值$V(x)$为分子除以分母,得到插值点$x$处的估计值$V(x)$。

[0049] 步骤S20、将得到的多源数据转化为文本数据,根据语义知识库提取多源数据中的语义信息,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,并对语义表征向量进行融合,得到融合后的多源学习数据。

[0050] 在该步骤中,根据语义知识库提取多源数据中的语义信息时,基于现有的WordNet、DBpedia和ConceptNet知识库,选择和构建覆盖多源异构数据领域的语义知识库;应用Stanford NER在标准化处理后的多源数据上识别出命名实体,并将识别出的实体与语义知识库中的条目进行匹配,在识别出的实体上进行词义消歧,于语义知识库提取实体和概念之间的语义关系,构建和生成包含实体和关系的语义图谱。

[0051] 通过以上详细步骤,可以系统化地从多源数据中提取语义信息,生成准确且结构清晰的语义图谱。每个步骤都有明确的执行方式,并且通过权利要求书保证了方法的创新性和合法性。

[0052] 示例性的,将多源数据转化为文本数据时,可以将上述收集到的多源数据转换为结构化的文本表示:小明的出勤率是95%。他的课堂参与度很高。期中考试成绩是85分,期末考试成绩是88分。作业完成时间平均为40分钟,正确率为90%。他完成了80%的在线课程,在线测试结果是85%。小明每天晚上7点到9点学习,学习路径是先复习课堂笔记,然后完成作业,最后进行在线测试。

[0053] 在本实施例中,参见图2所示,投影到共享语义空间生成词向量矩阵,对词向量矩阵分解后得到语义表征向量,包括:

步骤S201、将提取语义信息后的文本数据输入预训练的词向量模型,生成对应的词向量;

步骤S202、将词向量映射到一个共享的语义空间,使用对抗训练对齐不同源的词向量;

步骤S203、将对齐后的词向量组合构建形成完整的词向量矩阵;

步骤S204、对词向量矩阵进行中心化处理,减去每一列的均值,应用奇异值分解将词向量矩阵分解为三个矩阵U、Σ、V^T,其中:$A = U\Sigma V^T$式中,A是一个m×n的词向量矩阵;U是一个 m×n的正交矩阵,U列向量是A的左奇异向量;Σ是一个 m×n的对角矩阵,Σ对角线上的元素是A的奇异值;V^T是一个 m×n的正交矩阵,V^T行向量是A的右奇异向量的转置;

步骤S205、选取前k个奇异值及其对应的奇异向量,得到低维语义表征向量;选取前k个奇异值后,公式表示为:$A_k \approx U_k \Sigma_k V_k^T$式中,$U_k$是U的前k列,大小m×k;$\Sigma_k$是Σ的前k个奇异值,对应一个k×k的对角矩阵;$V_k^T$是$V^T$的前k行,大小为k×n;步骤S206、使用分解后的U矩阵中的前k列,作为词的低维语义表征向量。

[0054] 其中,所述词向量模型为预训练的Word2Vec模型,通过使用Word2Vec在语料库上进行训练得到,对抗训练时,还包括设计的一个对抗训练框架,该对抗训练框架包括生成器和判别器,生成器负责将词向量转换到共享语义空间,判别器则评估词向量是否来自同一数据源,通过反复训练生成器和判别器,使不同源的词向量在共享语义空间中具有一致性。

[0055] 通过上述步骤,通过上述步骤,小明的多源数据被成功投影到一个共享的语义空间,并生成了低维的语义表征向量。这些表征向量可以进一步用于生成个性化作业,确保作业内容能够准确反映小明的学习情况和需求。可以将多源数据中的词向量对齐到共享语义空间,并通过奇异值分解方法生成低维语义表征向量,这一过程确保了不同来源的数据能够在统一的语义空间中进行比较和分析,提供了高效且准确的语义表征。

[0056] 步骤S30、根据融合后的多源学习数据构建学生画像,识别出每个学生的知识薄弱点和学习兴趣点。

[0057] 在该步骤中,参见图3所示,构建学生画像时,包括以下步骤:

步骤S301、构建包含知识水平、学习习惯、兴趣爱好多维特征的学生画像,将融合后的多源学习数据$\mathbf{Z}$分解成多维的特征子集,建立知识点映射矩阵:$M = \begin{bmatrix} m_{11} m_{12} \cdots m_{1j} \\ m_{21} m_{22} \cdots m_{2j} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ m_{i1} m_{i2} \cdots m_{ij} \end{bmatrix}$

式中,m_{ij}表示第i个学习活动涉及第j个知识点的程度;

步骤S302、进行知识点掌握度计算,计算每个学生对各知识点的掌握度:

$$K = Z_k \times M$$

[0058] 其中,K是学生对各知识点的掌握度矩阵,其中,Z_k为分解成的特征子集;步骤S303、薄弱点识别,设定一个掌握度阈值θ,识别出掌握度低于阈值的知识点:$W = \{k_i | k_i < \theta\}$式中,W是学生的知识薄弱点集合;

步骤S304、使用K-means对学习兴趣特征子集进行聚类,识别出不同的兴趣点,对聚类结果进行标注,识别出每个兴趣点的含义;

步骤S305、将上述计算结果整合,生成每个学生的详细画像报告。

[0059] 示例性的,建立知识点映射矩阵,表示不同学习活动对于不同知识点的涉及程度

时,假设知识点如下:知识点A(代数);

知识点B(几何);

知识点C(微积分)。

[0060] 知识点映射矩阵如下表示:

那么,计算学生小明对各知识点的掌握度时,掌握度为:

掌握度=课堂表现×0.8+考试成绩×0.9+作业完成情况×0.85+在线学习×0.8。

[0061] 若小明的计算结果为:知识点A 0.85;知识点B 0.75;知识点C 0.65,设定掌握度阈值为0.70,识别出掌握度低于阈值的知识点。对学习兴趣特征(课堂参与度和在线课程观看记录)进行K-means聚类,得到聚类结果如下:

聚类1:高参与度和高完成率;

聚类2:中等参与度和中等完成率;

假设小明被聚类到聚类1,代表小明对较高参与度和高完成率的内容感兴趣,那么,整合结果,生成的小明的详细画像报告为:

知识水平:较好(但在知识点C上较弱);

学习习惯:出勤率高,学习时间集中在晚上,学习路径有条理;

兴趣爱好:对高参与度和高完成率的内容感兴趣。

[0062] 通过上述流程,本发明可以为小明生成一份个性化的作业,既能巩固他的薄弱知识点,又能激发他的学习兴趣,不仅能提高学生的学习效果,还能让学习过程更有趣、更高效。

[0063] 步骤S40、以语义表征向量和包括知识薄弱点和学习兴趣点的学生画像为输入,以个性化作业推荐为标签,训练作业指导模型。

[0064] 该步骤中,训练作业指导模型时,将生成的语义表征向量与对应的语义标签配对,形成训练数据集模型进行训练;其中,作业语义表征向量作为输入,知识薄弱点和学习兴趣点对应的语义信息作为标签;选择Transformer模型并使用生成的训练数据集进行模型训练,得到作业指导模型,利用训练好的模型生成个性化作业指导。

[0065] 步骤S50、逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集。

[0066] 该步骤中,生成知识点架构树时,根据章节内容和提取的关键词,初步建立知识点架构树的层级结构,通过语义分析和关联规则挖掘,确定各知识点之间的关系,去除冗余节点,优化架构树结构,将知识点与资料数据库中的资源进行匹配,根据得到的知识点架构树筛选出与其对应的资源子数据集。

[0067] 步骤S60、基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中推送依据所述关键词数据筛选的资源数据,生成符合每个学生的个性化作业。

[0068] 示例性的,逐章节遍历学生端上传的在线点选内容,生成知识点架构树并筛选资源子数据集中,若识别点选内容的关键词数据时,假设在某章节中,小明上传了关于“微积分”的在线点选内容,这些内容包括:界定积分、微分、牛顿-莱布尼茨公式。通过自然语言处理提取出以下关键词:

(1)界定积分;

(2)微分;

(3)牛顿-莱布尼茨公式。

[0069] 生成知识点架构树时,结合章节内容和提取的关键词,初步建立知识点架构树的层级结构:

微积分

├── 界定积分

├── 微分

└── 牛顿-莱布尼茨公式

然后进行语义分析和关联规则挖掘,通过语义分析和关联规则挖掘,确定各知识点之间的关系,得到:

界定积分与微分密切相关,可以通过牛顿-莱布尼茨公式联系在一起。

[0070] 然后,优化后的知识点架构树如下:

微积分

├── 界定积分

│ └── 牛顿-莱布尼茨公式

└── 微分

└── 牛顿-莱布尼茨公式

随后将知识点与资料数据库中的资源进行匹配,匹配结果为:

(1)界定积分:匹配到一些习题集、讲解视频等资源。

[0071] (2)微分:匹配到相关的在线课程、练习题等资源。

[0072] (3)牛顿-莱布尼茨公式:匹配到详细的理论讲解和应用实例等资源。

[0073] 最后,根据得到的知识点架构树筛选出与其对应的资源子数据集,资源子数据集为:

├── 界定积分

│ ├── 习题集

│ └── 讲解视频

├── 微分

│ ├── 在线课程

│ └── 练习题

└── 牛顿-莱布尼茨公式

├── 理论讲解

└── 应用实例

最后,根据小明的学生画像:

知识薄弱点:知识点C (知识点C包含界定积分和微分);

学习兴趣点:高参与度内容。

[0074] 基于作业指导模型和学生画像推送资源,生成个性化作业,从资源子数据集中选取适合小明的资源,对于小明的知识薄弱点(界定积分和微分),可以推送:界定积分的习题集和讲解视频,以及微分的在线课程和练习题。同时,考虑到小明的学习兴趣,可以选择比较互动性强的资源,如有趣的讲解视频和互动性的在线课程。最终生成的个性化作业如下:

1.界定积分:

·习题集:完成第3章第1-10题。

[0075] ·讲解视频:观看“界定积分的应用”视频并回答相关问题。

[0076] 2.微分:

·在线课程:完成第2节“微分的基本概念”并参与在线测验。

[0077] ·练习题:完成第4章练习题第1-15题。

[0078] 3.牛顿-莱布尼茨公式:

·理论讲解:阅读“牛顿-莱布尼茨公式及其应用”的讲义。

[0079] ·应用实例:完成相关的应用实例题,并提交报告。

[0080] 通过上述步骤,本发明为小明生成了一份详细且个性化的作业指导,确保小明能够针对自己的薄弱点进行有效的学习,同时保持对学习的高兴趣度。这种方法不仅提高了学习的针对性和有效性,还能激发学生的学习动力。

[0081] 通过上述步骤,本发明能够有效地利用多源数据融合生成符合学生个性化需求的作业,帮助学生更好地掌握知识点并提升学习兴趣和效果。

[0082] 本发明的一种基于多源数据融合的作业生成方法,相较于目前提出的在线学习平台和学习管理系统而言,本发明通过获取和融合多源异构数据,如学术成绩、课堂表现、作业成绩、在线学习记录等,全面覆盖学生的学习情况,有助于准确评估学生的学习状态,消除不同数据源之间的差异后,得到全面覆盖学生学习情况的多源数据。利用语义知识库提取多源数据中的语义信息,并投影到共享语义空间,生成词向量矩阵,通过分解和融合生成语义表征向量,确保数据中的重要信息能够被充分利用。基于融合后的多源学习数据,构建详细的学生画像,识别每个学生的知识薄弱点和学习兴趣点,提供了个性化学习的基础;以语义表征向量为输入,结合学生画像及其知识薄弱点和学习兴趣点,训练精确的作业指导模型,确保生成的作业能够针对学生的具体需求和兴趣。

[0083] 本发明的基于多源数据融合的作业生成方法,还能够通过逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据,生成对应的知识点架构树,确保作业内容与学生实际学习内容的高度相关性。基于生成的知识点架构树和学生画像,从资源子数据集中筛选出最适合的资源数据,确保学生获得的作业内容是最契合其学习需求的。通过作业指导模型和学生画像,生成符合每个学生的个性化作业,帮助学生更有针对性地进行学习,提升学习效果。

[0084] 应该理解的是,上述虽然是按照某一顺序描述的,但是这些步骤并不是必然按照上述顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,本实施例的一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0085] 在一个实施例中,本发明提供了一种基于多源数据融合的作业生成系统,用于执行上述基于多源数据融合的作业生成方法,该系统包括:

多源数据获取模块:用于获取多源异构数据,并对获取的多源异构数据进行标准化处理,生成全面覆盖学生学习情况的多源数据;

语义信息提取模块:用于将标准化处理后的多源数据转化为文本数据,根据语义

知识库,提取多源数据中的语义信息,并投影到共享语义空间,生成词向量矩阵;

词向量分解及融合模块:用于对词向量矩阵进行分解,得到语义表征向量,并对这些向量进行融合,生成融合后的多源学习数据;

学生画像构建模块:基于融合后的多源学习数据,构建学生画像,识别每个学生的知识薄弱点和学习兴趣点;

作业指导模型训练模块:用于以语义表征向量为输入,以对应的语义信息为标签,结合学生画像及其知识薄弱点和学习兴趣点,训练作业指导模型;

在线内容识别模块:用于逐章节遍历学生端上传的在线点选内容,识别点选内容的关键词数据;

架构树生成模块:用于根据识别出的关键词数据,生成对应的资料数据库的知识点架构树,并筛选出知识点架构树的资源子数据集;

作业生成模块:用于基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点,从资源子数据集中筛选资源数据,生成符合每个学生的个性化作业;

数据推送模块,用于将生成的个性化作业推送至学生端,确保每个学生都能收到符合其学习需求和兴趣的作业。

[0086] 本发明的基于多源数据融合的作业生成系统,通过上述模块的协同工作,能够高效、准确地生成个性化的学习作业,帮助学生更好地掌握知识,提升学习效果。

[0087] 在本实施例中,基于多源数据融合的作业生成系统在执行时采用如前述的一种基于多源数据融合的作业生成方法的步骤,因此,本实施例中对基于多源数据融合的作业生成系统的运行过程不再详细介绍。

[0088] 综上所述,本发明的基于多源数据融合的作业生成方法与系统通过全面的数据获取、智能的语义分析、精准的作业指导和高效的个性化作业生成,显著提升了作业生成的准确性和个性化程度,促进了学生的学习效果和教育质量的提升。

[0089] 在一个实施例中,在本发明的实施例中还提供了一种计算机设备,包括至少一个处理器,以及与所述至少一个处理器通信连接的存储器,所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器执行所述的基于多源数据融合的作业生成方法的步骤。

[0090] 在一个实施例中,本发明还提供了一种计算机可读存储介质,计算机可读存储介质存储有计算机指令,所述计算机指令用于使所述计算机执行所述的基于多源数据融合的作业生成方法的步骤。

[0091] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机指令表征的计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。

[0092] 非易失性存储器可包括只读存储器、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器或外部高速缓冲存储器。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器或动态随机存取存储器等。

[0093] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精

神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

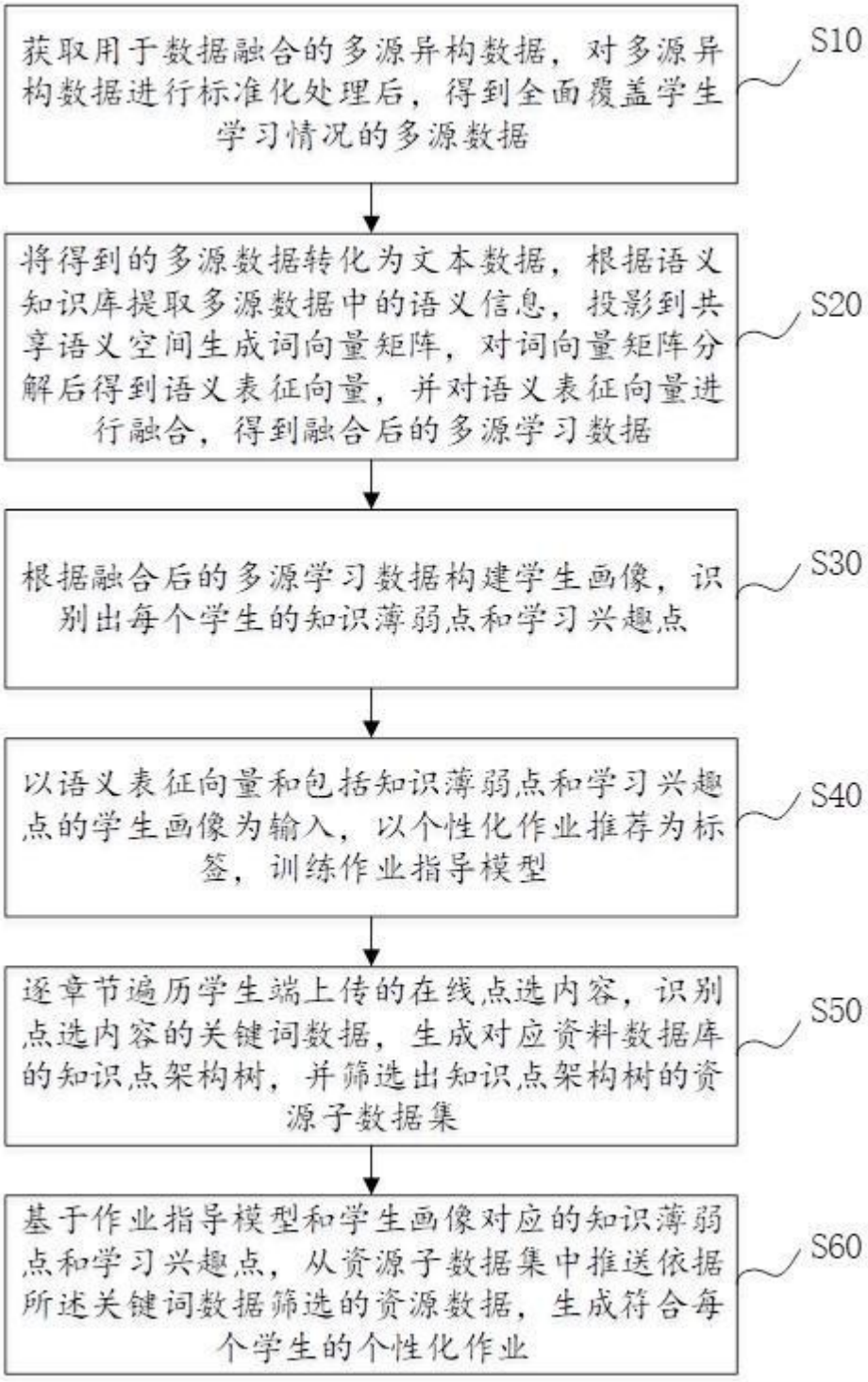
获取用于数据融合的多源异构数据，对多源异构数据进行标准化处理后，得到全面覆盖学生学习情况的多源数据
S10
将得到的多源数据转化为文本数据，根据语义知识库提取多源数据中的语义信息，投影到共享语义空间生成词向量矩阵，对词向量矩阵分解后得到语义表征向量，并对语义表征向量进行融合，得到融合后的多源学习数据
S20
根据融合后的多源学习数据构建学生画像，识别出每个学生的知识薄弱点和学习兴趣点
S30
以语义表征向量和包括知识薄弱点和学习兴趣点的学生画像为输入，以个性化作业推荐为标签，训练作业指导模型
S40
逐章节遍历学生端上传的在线点选内容，识别点选内容的关键词数据，生成对应资料数据库的知识点架构树，并筛选出知识点架构树的资源子数据集
S50
基于作业指导模型和学生画像对应的知识薄弱点和学习兴趣点，从资源子数据集中推送依据所述关键词数据筛选的资源数据，生成符合每个学生的个性化作业
S60

图 1

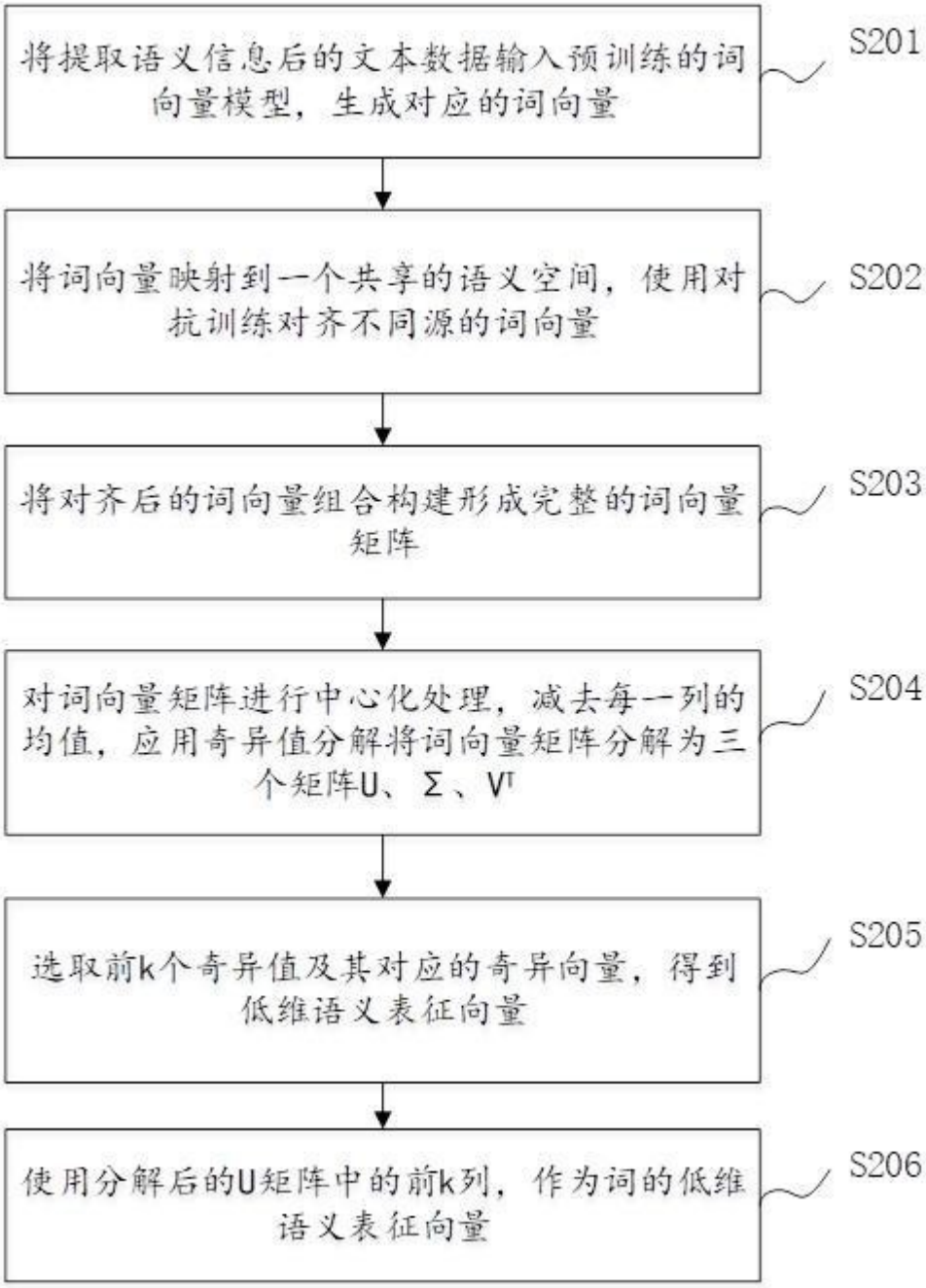
将提取语义信息后的文本数据输入预训练的词向量模型，生成对应的词向量
S201
将词向量映射到一个共享的语义空间，使用对抗训练对齐不同源的词向量
S202
将对齐后的词向量组合构建形成完整的词向量矩阵
S203
对词向量矩阵进行中心化处理，减去每一列的均值，应用奇异值分解将词向量矩阵分解为三个矩阵U、Σ、V^T
S204
选取前k个奇异值及其对应的奇异向量，得到低维语义表征向量
S205
使用分解后的U矩阵中的前k列，作为词的低维语义表征向量
S206

图 2

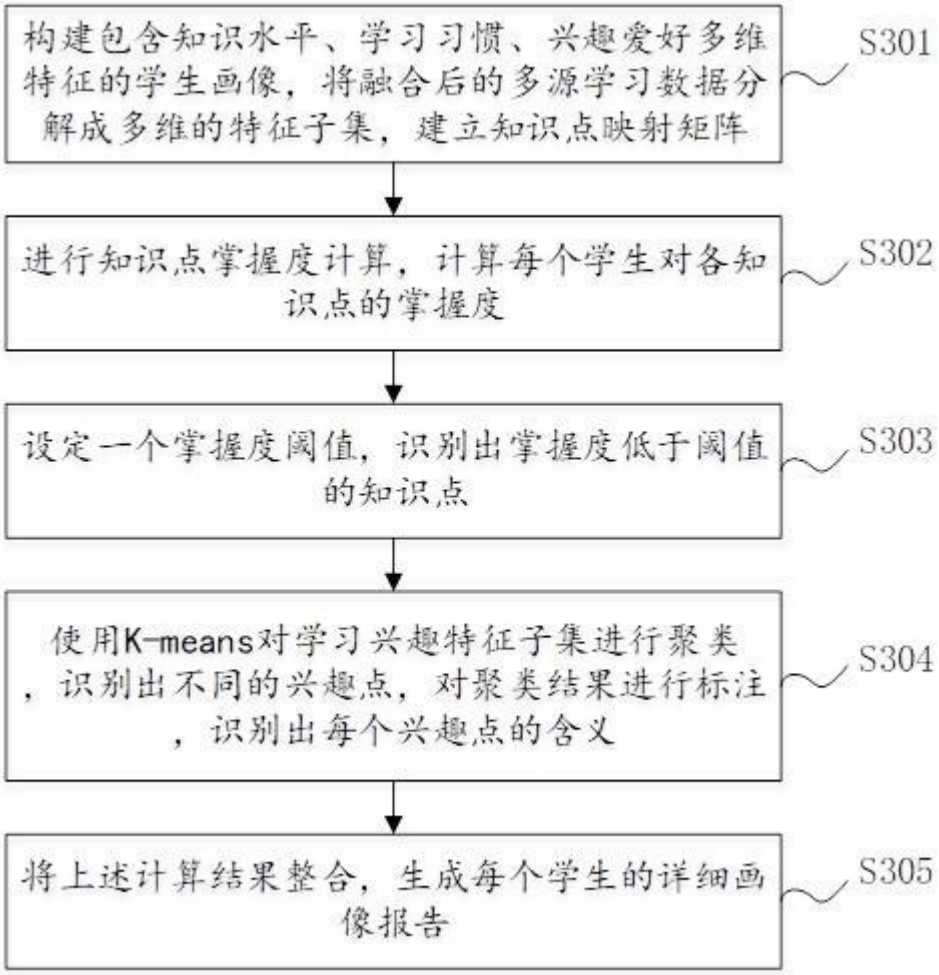
构建包含知识水平、学习习惯、兴趣爱好多维特征的学生画像，将融合后的多源学习数据分解成多维的特征子集，建立知识点映射矩阵
S301
进行知识点掌握度计算，计算每个学生对各知识点的掌握度
S302
设定一个掌握度阈值，识别出掌握度低于阈值的知识点
S303
使用K-means对学习兴趣特征子集进行聚类，识别出不同的兴趣点，对聚类结果进行标注，识别出每个兴趣点的含义
S304
将上述计算结果整合，生成每个学生的详细画像报告
S305

图 3