



(12) 发明专利申请

(10) 申请公布号 CN 117290596 A

(43) 申请公布日 2023. 12. 26

(21) 申请号 202311220669.7

(22) 申请日 2023.09.20

(71) 申请人 北京约来健康科技有限公司  
地址 100000 北京市昌平区回龙观镇回龙  
观西大街18号1段3层1-110

(72) 发明人 张航

(74) 专利代理机构 成都顶峰专利事务所(普通  
合伙) 51224  
专利代理师 叶昌威

(51) Int. Cl.

G06F 16/9535 (2019.01)

G06F 18/213 (2023.01)

G06F 18/25 (2023.01)

G06N 3/0442 (2023.01)

G06N 3/08 (2023.01)

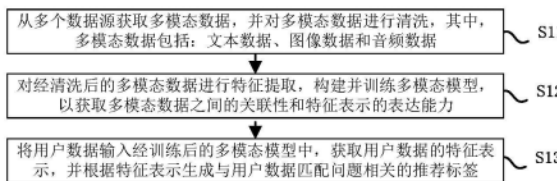
权利要求书2页 说明书10页 附图1页

(54) 发明名称

用于多模态数据模型的推荐标签生成方法、  
装置、设备及介质

(57) 摘要

本发明公开了一种用于多模态数据模型的推荐标签生成方法、装置、设备及介质,涉及信息技术领域。方法包括:从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能;将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。以此解决了现有技术中内外网数据交换存在着安全风险,传统的解决方案存在一些局限性的问题。



1. 一种用于多模态数据模型的推荐标签生成方法,其特征在于,方法包括:

从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,所述多模态数据包括:文本数据、图像数据和音频数据;

对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取所述多模态数据之间的关联性和特征表示的表达能力;

将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据所述特征表示生成与所述用户数据匹配问题相关的推荐标签。

2. 根据权利要求1所述的用于多模态数据模型的推荐标签生成方法,其特征在于,根据所述特征表示生成与匹配问题相关的推荐标签的步骤之后,还包括:

获取用户的选定结果,并根据所述选定结果评价推荐标签的准确度;

根据所述准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

3. 根据权利要求2所述的用于多模态数据模型的推荐标签生成方法,其特征在于,在获取用户的选定结果,并根据所述选定结果评价推荐标签的准确度的步骤之前,还包括:

根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序。

4. 根据权利要求3所述的用于多模态数据模型的推荐标签生成方法,其特征在于,根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序的步骤,包括:

将用户的历史选定行为特征和当前推荐结果,按顺序输入长短时记忆网络LSTM序列模型,建立用户行为特征序列和推荐标签序列之间的关系;

通过学习用户行为特征和推荐结果之间的时序依赖关系,对标签进行筛选和排序。

5. 根据权利要求1所述的用于多模态数据模型的推荐标签生成方法,其特征在于,对所述多模态数据进行特征提取的步骤,包括:

将文本数据转换为向量表示;

利用卷积神经网络进行图像特征提取;

从音频数据中提取音频特征,其中所述音频特征包括:声谱图或梅尔频率倒谱系数。

6. 一种用于多模态数据模型的推荐标签生成装置,其特征在于,包括:

数据获取模块,用于从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,所述多模态数据包括:文本数据、图像数据和音频数据;

模型训练模块,用于对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取所述多模态数据之间的关联性和特征表示的表达能力;

标签推荐模块,用于将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据所述特征表示生成与所述用户数据匹配问题相关的推荐标签。

7. 根据权利要求6所述的用于多模态数据模型的推荐标签生成装置,其特征在于,还包括准确度评价模块和增量训练模块,其中,

准确度评价模块,用于获取用户的选定结果,并根据所述选定结果评价推荐标签的准确度;

增量训练模块,用于根据所述准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

8. 根据权利要求7所述的用于多模态数据模型的推荐标签生成装置, 其特征在于, 还包括标签优化模块,

所述标签优化模块, 用于根据用户的历史选定行为特征和当前模型的推荐结果, 对推荐标签进行筛选和排序。

9. 一种计算机设备, 其特征在于, 包括有通信相连的存储器和处理器, 其中, 所述存储器用于存储计算机程序, 所述处理器用于读取所述计算机程序, 执行如权利要求1~5任一项中所述的用于多模态数据模型的推荐标签生成方法。

10. 一种计算机可读存储介质, 其特征在于, 所述计算机可读存储介质上存储有指令, 当所述指令在计算机上运行时, 执行如权利要求1~5任一项中所述的用于多模态数据模型的推荐标签生成方法。

## 用于多模态数据模型的推荐标签生成方法、装置、设备及介质

### 技术领域

[0001] 本发明涉及信息技术领域,尤其涉及一种用于多模态数据模型的推荐标签生成方法、装置、设备及介质。

### 背景技术

[0002] 多模态模型是一种结合文本与其它模态数据(如图像、音频等)进行处理和理解的核心技术。通过从多个模态输入中获取信息,这些模型可以更全面且准确地回答用户的问题。然而,目前这些模型在处理专业词汇和表达方面仍存在困难。专业知识和领域特定术语通常具有复杂语义含义,因此许多用户可能在提问时无法准确使用这些术语,从而影响模型对问题的理解和答案的准确性。准确的词汇和表达对于问题的精准度非常重要,特别是在专业领域如医学领域等问题中,其中涉及的术语和描述往往更加复杂。

[0003] 如果用户无法准确使用专业词汇和表达,多模态模型很难准确理解用户的问题并给出准确的回答。因此,问题的精准度对于回答的准确性至关重要。提问的精准度包括对于专业词汇和表达的理解等方面,这也导致那些专业度较低的用户很难从中获得高质量的推荐内容。这可能导致用户的困惑,并降低他们对多模态模型的使用意愿。对于专业度较低的用户来说,他们可能更需要准确的建议和指导,以帮助他们更好地理解 and 解决问题。因此,需要改进现有技术,提出更合理的技术方案来解决这些问题。

### 发明内容

[0004] 为了解决在现有技术中多模态模型对提问的精准度要求较高,导致那些专业度较低的用户很难从中获得高质量的推荐内容的问题。

[0005] 第一方面的,本发明实施例提出了一种用于多模态数据模型的推荐标签生成方法,方法包括:

[0006] 从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;

[0007] 对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;

[0008] 将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。

[0009] 优选地,并根据特征表示生成与匹配问题相关的推荐标签的步骤之后,还包括:

[0010] 获取用户的选定结果,并根据选定结果评价推荐标签的准确度;

[0011] 根据准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

[0012] 优选地,在获取用户的选定结果,并根据选定结果评价推荐标签的准确度的步骤之前,还包括:

[0013] 根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排

序。

[0014] 优选地,根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序的步骤,包括:

[0015] 将用户的历史选定行为特征和当前推荐结果,按顺序输入长短时记忆网络LSTM序列模型,建立用户行为特征序列和推荐标签序列之间的关系;

[0016] 通过学习用户行为特征和推荐结果之间的时序依赖关系,对标签进行筛选和排序。

[0017] 优选地,对多模态数据进行特征提取的步骤,包括:

[0018] 将文本数据转换为向量表示;

[0019] 利用卷积神经网络进行图像特征提取;

[0020] 从音频数据中提取音频特征,其中音频特征包括:声谱图或梅尔频率倒谱系数。

[0021] 第二方面的,本发明实施例提出了一种用于多模态数据模型的推荐标签生成装置,包括:

[0022] 数据获取模块,用于从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;

[0023] 模型训练模块,用于对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;

[0024] 标签推荐模块,用于将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。

[0025] 优选地,还包括准确度评价模块和增量训练模块,其中,

[0026] 准确度评价模块,用于获取用户的选定结果,并根据选定结果评价推荐标签的准确度;

[0027] 增量训练模块,用于根据准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

[0028] 优选地,还包括标签优化模块,

[0029] 标签优化模块,用于根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序。

[0030] 第三方面的,本发明实施例提出了一种计算机设备,包括有通信相连的存储器和处理器,其中,存储器用于存储计算机程序,处理器用于读取计算机程序,执行如上述实施例所提出的用于多模态数据模型的推荐标签生成方法。

[0031] 第四方面的,本发明实施例提出了一种计算机可读存储介质,计算机可读存储介质上存储有指令,当指令在计算机上运行时,执行如上述实施例所提出的用于多模态数据模型的推荐标签生成方法。

[0032] 有益效果:通过从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。通过引入一个标签推荐方法来改善多模态模型的用户体验和准确性,这个方法可以基于模型对问题和内容的理解,自动推荐相关的标签或关键

词。通过引导用户使用合适的标签,可以提高问题的精准度和模型的答案准确性。同时,这也能够帮助专业度较低的用户获得更高品质的推荐内容,提升他们的问题解决能力和满意度。以此解决了在现有技术中多模态模型对提问的精准度要求较高,导致那些专业度较低的用户很难从中获得高质量的推荐内容的问题。

### 附图说明

[0033] 通过阅读下文优选实施方式的详细描述,各种其它的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0034] 图1为本发明实施例提出的一种用于多模态数据模型的推荐标签生成方法的流程图;

[0035] 图2为本发明实施例提出的一种用于多模态数据模型的推荐标签生成装置的功能模块示意图;

[0036] 图3为本发明实施例提出的另一种用于多模态数据模型的推荐标签生成装置的功能模块示意图。

### 具体实施方式

[0037] 为了更清楚地说明发明实施例或现有技术中的技术方案,下面将结合附图和实施例或现有技术的描述对发明作简单地介绍,显而易见地,下面关于附图结构的描述仅仅是发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。在此需要说明的是,对于这些实施例方式的说明用于帮助理解发明,但并不构成对发明的限定。

[0038] 第一方面的,请参阅图1,为本发明实施例提出的一种用于多模态数据模型的推荐标签生成方法,应用于用于多模态数据模型的推荐标签生成装置;其中,用于多模态数据模型的推荐标签生成装置可以但不限于由具有一定计算资源的计算机设备执行,例如由个人计算机(Personal Computer,PC,指一种大小、价格和性能适用于个人使用的多用途计算机;台式机、笔记本电脑到小型笔记本电脑和平板电脑以及超级本等都属于个人计算机)、智能手机、个人数字助理(Personal digital assistant,PAD)或平台服务端等电子设备。存储端包括主板、存储器及外置数据交换模块,外置数据交换模块和存储器分别与主板连接,存储器内设置有加解密单元和身份数据只读存储单元,方法包括:从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。通过引入一个标签推荐方法来改善多模态模型的用户体验和准确性,这个方法可以基于模型对问题和内容的理解,自动推荐相关的标签或关键词。通过引导用户使用合适的标签,可以提高问题的精准度和模型的答案准确性。同时,这也能够帮助专业度较低的用户获得更高品质的推荐内容,提升他们的问题解决能力和满意度。以此解决了在现有技术中多模态模型对提问的精准度要求较高,导致那些专业度较低的用户很难从中获得高质量的推荐内容的问题。

[0039] 可以理解的,前述执行主体并不构成对本申请实施例的限定,相应的,本方法的运行步骤可以但不限于如下述步骤S11~S13所示。

[0040] 步骤S11,从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;

[0041] 多模态数据包括文本数据、图像数据和音频数据,文本数据源可以是来自网页、社交媒体、新闻文章、博客、论坛等文本内容的数据源,其中,这些文本数据可以包含用户评论、产品描述及新闻报道;图像数据可以是来自照片、图像数据库、社交媒体平台、视频帧等的图像数据源,其中,这些图像数据可以包含产品图片、用户上传的图片及场景照片;音频数据可以是来自录音、音乐库、声音数据库、视频音轨等的音频数据源,其中,这些音频数据可以包含语音讲解、音乐及声音效果。在获取到多模态数据后,对多模态数据进行数据清洗,包括检测并删除重复的数据样本,避免数据重复造成的偏差;检测并处理缺失的数据,如通过插值、填充默认值等方法来处理缺失数据;检测并处理异常的数据,如通过统计方法或者使用异常检测算法来识别和处理异常值;将不同数据源的数据格式进行标准化,确保数据格式的一致性,方便后续处理和分析。在对数据处理后,分别对文本数据、图像数据和音频数据进行清洗,包括:文本数据清洗:对文本数据进行去噪、去除特殊符号、拼写纠错、词干化等操作,以减少噪音和不必要的干扰因素;图像数据清洗:进行图像压缩、去除噪声、裁剪、调整亮度和对比度等操作,以提高图像质量和减少干扰;音频数据清洗:进行降噪处理、音频分割、提取音频特征等操作,以减少噪音和提取有效信息。清洗多模态数据可以提高数据的质量和准确性,为后续的特征提取和模型训练奠定基础。

[0042] 步骤S12,对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;

[0043] 分别对清洗后的文本数据、图像数据和音频数据进行特征提取,其中,对于文本数据,可以使用自然语言处理(NLP)技术,如词嵌入(Word Embedding)、TF-IDF等方法来提取文本的语义和统计特征。对于图像数据,可以使用计算机视觉技术,如卷积神经网络(CNN)进行图像特征提取,可以使用预训练的CNN模型(如VGG、ResNet、Inception等)或自定义的CNN模型。对于音频数据,可以使用音频处理技术,如梅尔频率倒谱系数(MFCC)来提取音频的声音特征。在特征提取后,构建多模态模型,将文本数据、图像数据和音频数据的特征表示进行融合;如,使用融合策略,如拼接、加权平均等,也可以使用多模态融合方法,如多层神经网络、注意力机制等。通过多模态特征提取和特征融合的方式,可以将不同数据源提供的信息整合起来,学习到数据之间的关联性,能够有效地捕捉多模态数据的特征信息,并以一种可解释的方式对其进行表达;通过训练多模态模型,模型可以学习到每个数据源的重要特征,以及特征之间的交互关系。其中,表达能力为层次性,以便于建立起特征表示的层次准确确定,通过多模态模型训练实现。

[0044] 步骤S13,将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。

[0045] 将用户的数据输入到经过训练的多模态模型中,该多模态模型已经学习了文本数据、图像数据和音频数据之间的关联性和特征表示的表达能力。通过将用户的数据输入到模型中,模型将会生成用户数据的特征表示。其中,特征表示是模型通过学习到的特征信息来表示用户数据的一种形式,这些特征可以捕捉到用户数据中的重要特征以及特征之间的

交互关系。基于用户数据的特征表示,可以进一步生成与用户数据匹配问题相关的推荐标签。推荐标签是对用户数据的描述或者标记,可以帮助对用户数据进行分类、识别或推荐。这些标签可以是关键词、主题、分类类别等,用于进一步处理用户数据或者提供给用户相关的推荐信息。通过将用户数据输入到多模态模型中,获取用户数据的特征表示,并利用这些特征表示生成与用户数据匹配问题相关的推荐标签,从而提供更准确和个性化的推荐标签服务。

[0046] 本实施例中,通过引入一个标签推荐方法来改善多模态模型的用户体验和准确性,这个方法可以基于模型对问题和内容的理解,自动推荐相关的标签或关键词。通过引导用户使用合适的标签,可以提高问题的精准度和模型的答案准确性。同时,这也能够帮助专业度较低的用户获得更高品质的推荐内容,提升他们的问题解决能力和满意度。以此解决了在现有技术中多模态模型对提问的精准度要求较高,导致那些专业度较低的用户很难从中获得高质量的推荐内容的问题。

[0047] 优选地,并根据特征表示生成与匹配问题相关的推荐标签的步骤之后,还包括:

[0048] 获取用户的选定结果,并根据选定结果评价推荐标签的准确度;

[0049] 在推荐标签给用户之后,用户将会选择其中一些标签作为他们认为匹配问题的正确标签,这些选定的结果将作为反馈;并根据选定结果评价推荐标签的准确度,如果用户选定的结果与生成的标签匹配,那么说明推荐标签的准确度较高。反之,则说明存在一定的误差或改进的空间。

[0050] 根据准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

[0051] 根据用户选定结果评价推荐标签的准确度,可以根据评价结果来调整多模态模型的超参数(如学习率、正则化参数等)、网络结构(如层数、神经元数等)和损失函数(如交叉熵损失、均方误差损失等)。通过这样的调整,可以优化模型的性能,并提高推荐标签的准确度。

[0052] 随着新增的多模态数据的更新,可以利用这些数据对模型进行增量训练,新增数据可以包括用户的选定结果和其它未使用过的多模态数据;通过使用这些数据来训练模型,可以进一步提升模型的性能和准确度。

[0053] 可以不断地根据用户的反馈来优化多模态模型和推荐标签的准确度,这将能够逐渐学习到用户的偏好和需求,并提供更加精准和个性化的推荐标签。

[0054] 优选地,在获取用户的选定结果,并根据选定结果评价推荐标签的准确度的步骤之前,还包括:

[0055] 根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序。

[0056] 根据用户过去的选择行为,例如之前选定的推荐标签、喜欢或不喜欢的内容等;通过分析这些历史选定行为特征,可以了解用户的偏好和兴趣。根据当前的多模态模型和用户的数据生成推荐标签,基于用户的历史选定行为特征和当前模型的推荐结果,会对生成的推荐标签进行筛选和排序的操作。筛选的目的是去除那些用户历史上不喜欢或选择次数较少的标签,以提高推荐的准确性和个性化。排序的目的是根据推荐标签的相关度、热门程度或其它因素,将最相关和最有价值的标签排在前面,以使用户更容易选择。



[0057] 能够利用用户的历史选定行为和当前模型的推荐结果,对推荐标签进行筛选和排序,从而提供更加符合用户兴趣和需求的推荐结果。这样的个性化筛选和排序可以帮助用户更快速地找到和选择匹配问题的正确标签。

[0058] 优选地,根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序的步骤,包括:

[0059] 将用户的历史选定行为特征和当前推荐结果,按顺序输入长短时记忆网络LSTM序列模型,建立用户行为特征序列和推荐标签序列之间的关系;

[0060] 将用户的历史选定行为特征和当前的推荐标签结果按照一定的顺序输入到LSTM序列模型中,LSTM是一种递归神经网络,能够对按时间顺序排列的输入数据进行处理,并捕捉序列数据中的时序依赖关系。通过LSTM序列模型,能够学习和建立用户历史选定行为特征序列和当前推荐结果序列之间的关系。LSTM会基于输入的序列数据,包括用户过去的选定行为特征和当前的推荐标签结果,自动学习它们之间的时序依赖关系,从而捕捉到用户在不同时间点上的兴趣和偏好变化。

[0061] 通过学习用户行为特征和推荐结果之间的时序依赖关系,对标签进行筛选和排序。

[0062] 基于LSTM序列模型学习到的用户行为特征和推荐结果之间的时序依赖关系,能够对推荐标签进行筛选和排序;这意味着可以利用学到的模式和规律,根据用户的历史行为和当前推荐结果对标签进行优化,将最相关、有价值的标签放在前面,同时过滤掉与用户兴趣不相关的标签。

[0063] 能够基于用户的历史选定行为特征和当前推荐结果,利用LSTM序列模型来学习用户行为特征和推荐结果之间的时序依赖关系,并对推荐标签进行筛选和排序。这样的个性化筛选和排序能够更好地满足用户的兴趣和需求,提供更准确和个性化的推荐标签。

[0064] 优选地,对多模态数据进行特征提取的步骤,包括:

[0065] 将文本数据转换为向量表示;

[0066] 为了将文本数据用于特征提取,会将其转换为向量表示。这可以通过使用自然语言处理(NLP)技术来实现,比如词嵌入(Word Embedding)方法。词嵌入可以将每个单词映射到一个固定长度的向量,以捕捉单词之间的语义关系和上下文信息。

[0067] 利用卷积神经网络进行图像特征提取;

[0068] 对于图像数据,可以使用卷积神经网络(CNN)进行特征提取。CNN在图像识别方面表现出色,能够通过卷积和池化等操作捕捉图像的局部和全局特征。通过在预训练的CNN模型(如VGG、ResNet、Inception等)上进行特征提取,可以获得图像数据的高级特征表示。

[0069] 从音频数据中提取音频特征,其中音频特征包括:声谱图或梅尔频率倒谱系数。

[0070] 对于音频数据,需要从中提取出有用的音频特征。常用的音频特征之一是声谱图(Spectrogram),它显示了声音在时间和频率上的变化。另一个常用的音频特征是梅尔频率倒谱系数(MFCC),它是一种在语音和音频处理中常用的特征表示方法,可以提取声音的频谱轮廓和共振峰。

[0071] 文本数据将被转换为向量表示,图像数据利用CNN进行特征提取,音频数据则提取声谱图或梅尔频率倒谱系数等音频特征。这样的特征提取过程有助于将多模态数据转化为计算机可以理解和处理特征表示形式,从而为后续模型训练和推荐等任务提供基础。

[0072] 第二方面的,本发明实施例提出了一种用于多模态数据模型的推荐标签生成装置100,包括:

[0073] 数据获取模块110,用于从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;

[0074] 多模态数据包括文本数据、图像数据和音频数据,文本数据源可以是来自网页、社交媒体、新闻文章、博客、论坛等文本内容的数据源,其中,这些文本数据可以包含用户评论、产品描述及新闻报道;图像数据可以是来自照片、图像数据库、社交媒体平台、视频帧等的图像数据源,其中,这些图像数据可以包含产品图片、用户上传的图片及场景照片;音频数据可以是来自录音、音乐库、声音数据库、视频音轨等的音频数据源,其中,这些音频数据可以包含语音讲解、音乐及声音效果。在获取到多模态数据后,对多模态数据进行数据清洗,包括检测并删除重复的数据样本,避免数据重复造成的偏差;检测并处理缺失的数据,如通过插值、填充默认值等方法来处理缺失数据;检测并处理异常的数据,如通过统计方法或者使用异常检测算法来识别和处理异常值;将不同数据源的数据格式进行标准化,确保数据格式的一致性,方便后续处理和分析。在对数据处理后,分别对文本数据、图像数据和音频数据进行清洗,包括:文本数据清洗:对文本数据进行去噪、去除特殊符号、拼写纠错、词干化等操作,以减少噪音和不必要的干扰因素;图像数据清洗:进行图像压缩、去除噪声、裁剪、调整亮度和对比度等操作,以提高图像质量和减少干扰;音频数据清洗:进行降噪处理、音频分割、提取音频特征等操作,以减少噪音和提取有效信息。清洗多模态数据可以提高数据的质量和准确性,为后续的特征提取和模型训练奠定基础。

[0075] 模型训练模块120,用于对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;

[0076] 分别对清洗后的文本数据、图像数据和音频数据进行特征提取,其中,对于文本数据,可以使用自然语言处理(NLP)技术,如词嵌入(Word Embedding)、TF-IDF等方法来提取文本的语义和统计特征。对于图像数据,可以使用计算机视觉技术,如卷积神经网络(CNN)进行图像特征提取,可以使用预训练的CNN模型(如VGG、ResNet、Inception等)或自定义的CNN模型。对于音频数据,可以使用音频处理技术,如梅尔频率倒谱系数(MFCC)来提取音频的声音特征。在特征提取后,构建多模态模型,将文本数据、图像数据和音频数据的特征表示进行融合;如,使用融合策略,如拼接、加权平均等,也可以使用多模态融合方法,如多层神经网络、注意力机制等。通过多模态特征提取和特征融合的方式,可以将不同数据源提供的信息整合起来,学习到数据之间的关联性,能够有效地捕捉多模态数据的特征信息,并以一种可解释的方式对其进行表达;通过训练多模态模型,模型可以学习到每个数据源的重要特征,以及特征之间的交互关系。

[0077] 标签推荐模块130,用于将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。

[0078] 将用户的数据输入到经过训练的多模态模型中,该多模态模型已经学习了文本数据、图像数据和音频数据之间的关联性和特征表示的表达能力。通过将用户的数据输入到模型中,模型将会生成用户数据的特征表示。其中,特征表示是模型通过学习到的特征信息来表示用户数据的一种形式,这些特征可以捕捉到用户数据中的重要特征以及特征之间的交互关系。基于用户数据的特征表示,可以进一步生成与用户数据匹配问题相关的推荐标

签。推荐标签是对用户数据的描述或者标记,可以帮助对用户数据进行分类、识别或推荐。这些标签可以是关键词、主题、分类类别等,用于进一步处理用户数据或者提供给用户相关的推荐信息。通过将用户数据输入到多模态模型中,获取用户数据的特征表示,并利用这些特征表示生成与用户数据匹配问题相关的推荐标签,从而提供更准确和个性化的推荐标签服务。

[0079] 本实施例中,通过引入一个标签推荐方法来改善多模态模型的用户体验和准确性,这个方法可以基于模型对问题和内容的理解,自动推荐相关的标签或关键词。通过引导用户使用合适的标签,可以提高问题的精准度和模型的答案准确性。同时,这也能够帮助专业度较低的用户获得更高品质的推荐内容,提升他们的问题解决能力和满意度。以此解决了在现有技术中多模态模型对提问的精准度要求较高,导致那些专业度较低的用户很难从中获得高质量的推荐内容的问题。

[0080] 优选地,还包括准确度评价模块150和增量训练模块160,其中,

[0081] 准确度评价模块150,用于获取用户的选定结果,并根据选定结果评价推荐标签的准确度;

[0082] 在推荐标签给用户之后,用户将会选择其中一些标签作为他们认为匹配问题的正确标签,这些选定的结果将作为反馈;并根据选定结果评价推荐标签的准确度,如果用户选定的结果与生成的标签匹配,那么说明推荐标签的准确度较高。反之,则说明存在一定的误差或改进的空间。

[0083] 增量训练模块160,用于根据准确度调整多模态模型的超参数、网络结构和损失函数,并根据新增的多模态数据,进行模型的增量训练。

[0084] 根据用户选定结果评价推荐标签的准确度,可以根据评价结果来调整多模态模型的超参数(如学习率、正则化参数等)、网络结构(如层数、神经元数等)和损失函数(如交叉熵损失、均方误差损失等)。通过这样的调整,可以优化模型的性能,并提高推荐标签的准确度。

[0085] 随着新增的多模态数据的更新,可以利用这些数据对模型进行增量训练,新增数据可以包括用户的选定结果和其它未使用过的多模态数据;通过使用这些数据来训练模型,可以进一步提升模型的性能和准确度。

[0086] 可以不断地根据用户的反馈来优化多模态模型和推荐标签的准确度,这将能够逐渐学习到用户的偏好和需求,并提供更加精准和个性化的推荐标签。

[0087] 优选地,还包括标签优化模块140,

[0088] 标签优化模块140,用于根据用户的历史选定行为特征和当前模型的推荐结果,对推荐标签进行筛选和排序。

[0089] 根据用户过去的选择行为,例如之前选定的推荐标签、喜欢或不喜欢的内容等;通过分析这些历史选定行为特征,可以了解用户的偏好和兴趣。根据当前的多模态模型和用户的数据生成推荐标签,基于用户的历史选定行为特征和当前模型的推荐结果,会对生成的推荐标签进行筛选和排序的操作。筛选的目的是去除那些用户历史上不喜欢或选择次数较少的标签,以提高推荐的准确性和个性化。排序的目的是根据推荐标签的相关度、热门程度或其它因素,将最相关和最有价值的标签排在前面,以使用户更容易选择。

[0090] 能够利用用户的历史选定行为和当前模型的推荐结果,对推荐标签进行筛选和排

序,从而提供更加符合用户兴趣和需求的推荐结果。这样的个性化筛选和排序可以帮助用户更快速找到和选择匹配问题的正确标签。

[0091] 第三方面的,还提出了一种用于多模态数据模型的推荐标签生成计算机设备,在硬件层面,包括:数据接口;存储器,用于存储指令;处理器,用于读取存储器中存储的指令,并根据指令执行实施例一中的用于多模态数据模型的推荐标签生成方法:步骤S11,从多个数据源获取多模态数据,并对多模态数据进行清洗,其中,多模态数据包括:文本数据、图像数据和音频数据;步骤S12,对经清洗后的多模态数据进行特征提取,构建并训练多模态模型,以获取多模态数据之间的关联性和特征表示的表达能力;步骤S13,将用户数据输入经训练后的多模态模型中,获取用户数据的特征表示,并根据特征表示生成与用户数据匹配问题相关的推荐标签。

[0092] 可选地,该设备还包括内部总线。处理器与存储器和数据接口可以通过内部总线相互连接,该内部总线可以是ISA(Industry Standard Architecture,工业标准体系结构)总线、PCI(Peripheral Component Interconnect,外设部件互连标准)总线或EISA(Extended Industry Standard Architecture,扩展工业标准结构)总线等。总线可以分为地址总线、数据总线、控制总线等。

[0093] 存储器可以但不限于包括随机存取存储器(Random Access Memory,RAM)、只读存储器(Read Only Memory,ROM)、闪存(Flash Memory)、先进先出存储器(First Input First Output,FIFO)和/或先进后出存储器(First In Last Out,FILO)等。处理器可以是通用处理器,包括中央处理器(Central Processing Unit,CPU)、网络处理器(Network Processor,NP)等;还可以是数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其它可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0094] 本发明的第四实施例提出的设备的工作过程、工作细节和技术效果,可以参见第一实施例,于此不再赘述。

[0095] 本发明的第四实施例提供了一种存储包含有第一实施例的用于多模态数据模型的推荐标签生成方法的指令的计算机可读存储介质,即计算机可读存储介质上存储有指令,当指令在计算机上运行时,执行如第一方面的用于多模态数据模型的推荐标签生成方法。其中,计算机可读存储介质是指存储数据的载体,可以但不限于包括软盘、光盘、硬盘、闪存、优盘和/或记忆棒(Memory Stick)等,计算机可以是通用计算机、专用计算机、计算机网络、或者其它可编程装置。

[0096] 本发明的第四实施例提出的计算机可读存储介质的工作过程、工作细节和技术效果,可以参见第一实施例,于此不再赘述。

[0097] 本实施例第五方面提供了一种包含指令的计算机程序产品,当指令在计算机上运行时,使计算机执行如第一实施例的用于多模态数据模型的推荐标签生成方法,其中,计算机可以是通用计算机、专用计算机、计算机网络、或者其它可编程装置。

[0098] 以上所描述的多个实施例仅仅是示意性的,其中作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的

部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实时。

[0099] 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各实施方式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。基于这样的理解,上述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,该计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台仓库代码的合并装置执行各个实施例或者实施例的某些部分的方法。

[0100] 最后应说明的是:以上仅为发明的优选实施例而已,并不用于限制发明的保护范围。凡在发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在发明的保护范围之内。

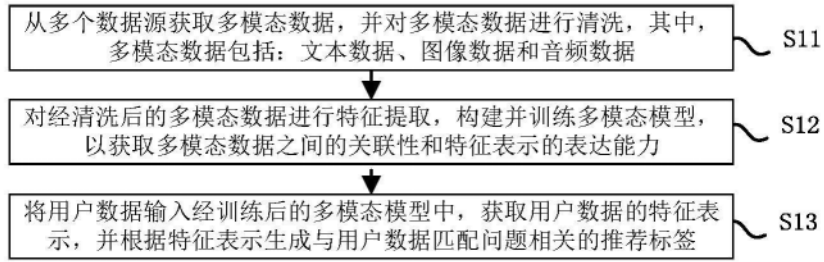


图1

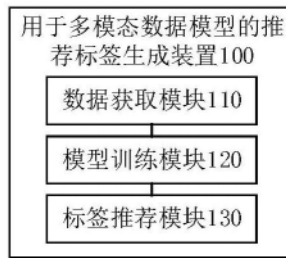


图2



图3