



(12) 发明专利

(10) 授权公告号 CN 114330475 B

(45) 授权公告日 2024.10.15

(21) 申请号 202111270814.3

G06N 3/08 (2023.01)

(22) 申请日 2021.10.29

G06V 10/42 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/74 (2022.01)

申请公布号 CN 114330475 A

G06V 10/771 (2022.01)

G06V 10/80 (2022.01)

(43) 申请公布日 2022.04.12

(56) 对比文件

(73) 专利权人 腾讯科技(深圳)有限公司

CN 109492101 A, 2019.03.19

地址 518000 广东省深圳市南山区高新区

CN 113076433 A, 2021.07.06

科技中一路腾讯大厦35层

审查员 胡晓雨

(72) 发明人 安涵

(74) 专利代理机构 北京市立方律师事务所

11330

专利代理师 张筱宁

(51) Int. Cl.

G06F 18/22 (2023.01)

G06N 3/0455 (2023.01)

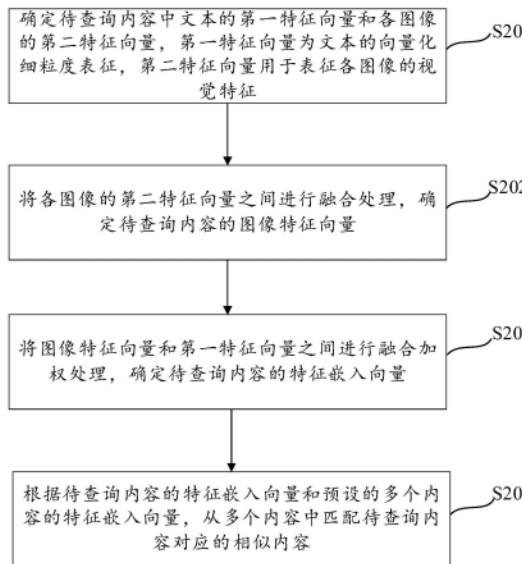
权利要求书3页 说明书17页 附图5页

(54) 发明名称

内容匹配方法、装置、设备、存储介质及计算机程序产品

(57) 摘要

本申请实施例提供了一种内容匹配方法、装置、设备、计算机可读存储介质及计算机程序产品,涉及人工智能领域,该方法包括:确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;根据待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容。该方法实现了更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。



1. 一种内容匹配方法,其特征在于,包括:

确定待查询内容中文本的第一特征向量和各图像的第二特征向量,所述第一特征向量为所述文本的向量化细粒度表征,所述第二特征向量用于表征所述各图像的视觉特征,所述文本的向量化细粒度表征用于表示基于聚类的词的粒度;

将所述各图像的第二特征向量之间进行融合处理,确定所述待查询内容的图像特征向量;

将所述图像特征向量和所述第一特征向量之间进行融合加权处理,确定所述待查询内容的特征嵌入向量;

根据所述待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从所述多个内容中匹配所述待查询内容对应的相似内容;

所述将所述图像特征向量和所述第一特征向量之间进行融合加权处理,确定所述待查询内容的特征嵌入向量,包括:

将所述图像特征向量和所述第一特征向量输入至神经网络模型的第二文本图像特征融合模型,进行融合处理,得到所述待查询内容的N维初始特征嵌入向量;

将所述待查询内容的N维初始特征嵌入向量输入至神经网络模型的加权拼接模型,将所述N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到所述待查询内容的特征嵌入向量,所述N为正整数。

2. 根据权利要求1所述的方法,其特征在于,所述确定所述待查询内容中文本的第一特征向量,包括:

将所述待查询内容中文本输入至神经网络模型的文本特征提取模型的匹配子模型,将所述文本和预设的多个聚类集合经过匹配处理,确定所述文本的第一特征嵌入向量对应的一个聚类集合,所述多个聚类集合包括所述一个聚类集合,多个聚类集合中的每个所述聚类集合包括多个归属于同一类型的文本标签;

根据所述文本的第一特征嵌入向量和所述聚类集合中各文本标签的特征嵌入向量,确定所述第一特征向量。

3. 根据权利要求2所述的方法,其特征在于,所述根据所述文本的第一特征嵌入向量和所述聚类集合中各文本标签的特征嵌入向量,确定所述第一特征向量,包括:

确定所述文本的第一特征嵌入向量和各文本标签的特征嵌入向量之间的第一相似度;

根据所述各文本标签的特征嵌入向量和各第一相似度对应的权重,确定所述第一特征向量,所述各第一相似度与所述各第一相似度对应的权重之间呈正相关。

4. 根据权利要求3所述的方法,其特征在于,所述根据所述各文本标签的特征嵌入向量和各第一相似度对应的权重,确定所述第一特征向量,包括:

将各第一相似度输入至神经网络模型的文本特征提取模型的排序子模型,将各第一相似度从大到小进行排序,确定排序在前的M个第一相似度;

根据所述M个第一相似度对应的权重,以及所述M个第一相似度对应的文本标签的特征嵌入向量,确定所述第一特征向量,M为正整数。

5. 根据权利要求2所述的方法,其特征在于,在所述确定所述待查询内容中文本的第一特征向量之前,还包括:

获取多个文本标签;

将所述多个文本标签进行聚类处理,得到所述多个聚类集合,所述多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签。

6. 根据权利要求1所述的方法,其特征在于,所述确定所述待查询内容中各图像的第二特征向量,包括:

将所述待查询内容中各图像分别输入至神经网络模型的图像特征提取模型的训练后的图像编码器,对所述各图像分别进行编码处理,得到所述待查询内容中各图像的第二特征向量。

7. 根据权利要求6所述的方法,其特征在于,在所述确定所述待查询内容中各图像的第二特征向量之前,还包括:

获取预设数据集中多个文本样本和多个图像样本;

将所述多个文本样本输入至所述图像特征提取模型的文本编码器,并将所述多个图像样本输入至所述图像特征提取模型的图像编码器,进行对比学习处理,得到多个正样本对,所述多个正样本对中的每个正样本对包括一个图像样本和一个文本样本,所述一个图像样本和一个文本样本之间的第二相似度大于或等于预设相似度阈值;

确定各正样本对的损失函数值,所述损失函数值用于表征所述各正样本对中图像样本的特征嵌入向量和文本样本的特征嵌入向量之间的距离;

基于所述各正样本对的损失函数值,更新所述文本编码器和所述图像编码器;

若所述各正样本对的损失函数值小于或等于预设损失函数值阈值,则结束所述文本编码器和所述图像编码器的训练,并将训练得到的图像编码器作为所述图像特征提取模型的训练后的图像编码器。

8. 根据权利要求1所述的方法,其特征在于,所述将所述各图像的第二特征向量之间进行融合处理,确定所述待查询内容的图像特征向量,包括:

将所述各图像的第二特征向量输入至神经网络模型的第一文本图像特征融合模型,进行融合处理,得到所述待查询内容的图像特征向量;

所述神经网络模型还包括文本特征提取模型、图像特征提取模型、第二文本图像特征融合模型和加权拼接模型。

9. 一种内容匹配装置,其特征在于,包括:

第一处理模块,用于确定待查询内容中文本的第一特征向量和各图像的第二特征向量,所述第一特征向量为所述文本的向量化细粒度表征,所述第二特征向量用于表征所述各图像的视觉特征,所述文本的向量化细粒度表征用于表示基于聚类的词的粒度;

第二处理模块,用于将所述各图像的第二特征向量之间进行融合处理,确定所述待查询内容的图像特征向量;

第三处理模块,用于将所述图像特征向量和所述第一特征向量之间进行融合加权处理,确定所述待查询内容的特征嵌入向量;

第四处理模块,用于根据所述待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从所述多个内容中匹配所述待查询内容对应的相似内容;

所述第三处理模块,具体用于:

将所述图像特征向量和所述第一特征向量输入至神经网络模型的第二文本图像特征融合模型,进行融合处理,得到所述待查询内容的N维初始特征嵌入向量;

将所述待查询内容的N维初始特征嵌入向量输入至神经网络模型的加权拼接模型,将所述N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到所述待查询内容的特征嵌入向量,所述N为正整数。

10. 一种电子设备,包括存储器、处理器及存储在存储器上的计算机程序,其特征在于,所述处理器执行所述计算机程序以实现权利要求1-8任一项所述方法的步骤。

11. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1-8任一项所述方法的步骤。

12. 一种计算机程序产品,包括计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1-8中任一项所述方法的步骤。

内容匹配方法、装置、设备、存储介质及计算机程序产品

技术领域

[0001] 本申请涉及计算机技术领域,具体而言,本申请涉及一种内容匹配方法、装置、设备、存储介质及计算机程序产品。

背景技术

[0002] 现有技术中内容匹配的应用场景可以是文章检索,文章检索广泛应用于文章排重等领域,通过文章检索,可以在数据库中查询是否存在与待查询文章相似的内容。现有技术中的文章检索方法直接使用TF-IDF(Term Frequency-Inverse Document Frequency,词频-逆向文件频率),TF-IDF是一种用于信息检索与数据挖掘的常用加权技术,TF-IDF将词向量加权以获得文章的文本特征向量,但是TF-IDF只考虑了文章的文本特征的向量化,从而导致文章的特征粒度粗以及信息缺失,文章检索的准确率不高,即内容匹配的准确率不高。

发明内容

[0003] 本申请针对现有的方式的缺点,提出一种内容匹配方法、装置、设备、计算机可读存储介质及计算机程序产品,用于解决如何提升内容匹配的准确率的问题。

[0004] 第一方面,本申请提供了一种内容匹配方法,包括:

[0005] 确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;

[0006] 将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;

[0007] 将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;

[0008] 根据待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容。

[0009] 在一个实施例中,确定待查询内容中文本的第一特征向量,包括:

[0010] 将待查询内容中文本输入至神经网络模型的文本特征提取模型的匹配子模型,将文本和预设的多个聚类集合经过匹配处理,确定文本的第一特征嵌入向量对应的一个聚类集合,多个聚类集合包括一个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签;

[0011] 根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量。

[0012] 在一个实施例中,根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量,包括:

[0013] 确定文本的第一特征嵌入向量和各文本标签的特征嵌入向量之间的第一相似度;

[0014] 根据各文本标签的特征嵌入向量和各第一相似度对应的权重,确定第一特征向量,各第一相似度与各第一相似度对应的权重之间呈正相关。

[0015] 在一个实施例中,根据各文本标签的特征嵌入向量和各第一相似度对应的权重,确定第一特征向量,包括:

[0016] 将各第一相似度输入至神经网络模型的文本特征提取模型的排序子模型,将各第一相似度从大到小进行排序,确定排序在前的M个第一相似度;

[0017] 根据M个第一相似度对应的权重,以及M个第一相似度对应的文本标签的特征嵌入向量,确定第一特征向量,M为正整数。

[0018] 在一个实施例中,在确定待查询内容中文本的第一特征向量之前,还包括:

[0019] 获取多个文本标签;

[0020] 将多个文本标签进行聚类处理,得到多个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签。

[0021] 在一个实施例中,确定待查询内容中各图像的第二特征向量,包括:

[0022] 将待查询内容中各图像分别输入至神经网络模型的图像特征提取模型的训练后的图像编码器,对各图像分别进行编码处理,得到待查询内容中各图像的第二特征向量。

[0023] 在一个实施例中,在确定待查询内容中各图像的第二特征向量之前,还包括:

[0024] 获取预设数据集中多个文本样本和多个图像样本;

[0025] 将多个文本样本输入至所述图像特征提取模型的文本编码器,并将多个图像样本输入至图像特征提取模型的图像编码器,进行对比学习处理,得到多个正样本对,多个正样本对中的每个正样本对包括一个图像样本和一个文本样本,一个图像样本和一个文本样本之间的第二相似度大于或等于预设相似度阈值;

[0026] 确定各正样本对的损失函数值,损失函数值用于表征各正样本对中图像样本的特征嵌入向量和文本样本的特征嵌入向量之间的距离;

[0027] 基于各正样本对的损失函数值,更新文本编码器和所述图像编码器;

[0028] 若各正样本对的损失函数值小于或等于预设损失函数值阈值,则结束文本编码器和图像编码器的训练,并将训练得到的图像编码器作为图像特征提取模型的训练后的图像编码器。

[0029] 在一个实施例中,将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量,包括:

[0030] 将各图像的第二特征向量输入至神经网络模型的第一文本图像特征融合模型,进行融合处理,得到待查询内容的图像特征向量;

[0031] 神经网络模型还包括文本特征提取模型、图像特征提取模型、第二文本图像特征融合模型和加权拼接模型。

[0032] 在一个实施例中,将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量,包括:

[0033] 将图像特征向量和第一特征向量输入至神经网络模型的第二文本图像特征融合模型,进行融合处理,得到待查询内容的N维初始特征嵌入向量;

[0034] 将待查询内容的N维初始特征嵌入向量输入至神经网络模型的加权拼接模型,将N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待查询内容的特征嵌入向量,N为正整数。

[0035] 第二方面,本申请提供了一种内容匹配装置,包括:

[0036] 第一处理模块,用于确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;

[0037] 第二处理模块,用于将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;

[0038] 第三处理模块,用于将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;

[0039] 第四处理模块,用于根据待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容。

[0040] 第三方面,本申请提供了一种电子设备,包括:处理器、存储器和总线;

[0041] 总线,用于连接处理器和存储器;

[0042] 存储器,用于存储操作指令;

[0043] 处理器,用于通过调用操作指令,执行本申请第一方面的内容匹配方法。

[0044] 第四方面,本申请提供了一种计算机可读存储介质,存储有计算机程序,计算机程序被用于执行本申请第一方面的内容匹配方法。

[0045] 第五方面,本申请提供了一种计算机程序产品,包括计算机程序,计算机程序被处理器执行时实现本申请第一方面中内容匹配方法的步骤。

[0046] 本申请实施例提供的技术方案,至少具有如下有益效果:

[0047] 确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;如此,将文本的第一特征向量和图像特征向量之间进行融合加权处理,实现了待查询内容的特征嵌入向量的提取,即待查询内容的多模态特征的提取,从而通过待查询内容的特征嵌入向量,更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

附图说明

[0048] 为了更清楚地说明本申请实施例中的技术方案,下面将对本申请实施例描述中所需要使用的附图作简单地介绍。

[0049] 图1为本申请实施例提供的内容匹配系统的架构示意图;

[0050] 图2为本申请实施例提供的一种内容匹配方法的流程示意图;

[0051] 图3为本申请实施例提供的一种原创文章相似召回的流程示意图;

[0052] 图4为本申请实施例提供的一种文章检索的示意图;

[0053] 图5为本申请实施例提供的一种内容匹配装置的结构示意图;

[0054] 图6为本申请实施例提供的一种电子设备的结构示意图。

具体实施方式

[0055] 下面结合本申请中的附图描述本申请的实施例。应理解,下面结合附图所阐述的实施方式,是用于解释本申请实施例的技术方案的示例性描述,对本申请实施例的技术方

案不构成限制。

[0056] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本申请实施例所使用的术语“包括”以及“包含”是指相应特征可以实现为所呈现的特征、信息、数据、步骤、操作、元件和/或组件,但不排除实现为本技术领域所支持其他特征、信息、数据、步骤、操作、元件、组件和/或它们的组合等。应该理解,当我们称一个元件被“连接”或“耦接”到另一元件时,该一个元件可以直接连接或耦接到另一元件,也可以指该一个元件和另一元件通过中间元件建立连接关系。此外,这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的术语“和/或”指示该术语所限定的项目中的至少一个,例如“A和/或B”指示实现为“A”,或者实现为“A”,或者实现为“A和B”。

[0057] 在相关技术中,TD-IDF加权的方法是根据TF-IDF权重,对文章中所有词向量进行加权求和。但是发明人研究发现至少存在以下问题:

[0058] (1) 直接使用TF-IDF将词向量加权以获得文本向量,没有考虑到词的顺序问题。

[0059] (2) 对于公开数据集学习到的特征提取器,该特征提取器提取的特征存在特征粒度粗的问题。

[0060] (3) TF-IDF只考虑了文章的文本特征的向量化,文章的特征向量缺失了对文章中图像信息的表征。

[0061] TD-IDF加权的方法直接使用TF-IDF将词向量加权以获得文本向量;该方案一方面没有考虑到词的顺序问题;另一方面,从公开数据集学习到的特征提取器,并不满足图文本标签任务对细粒度特征的要求。

[0062] 基于此,为了解决现有内容匹配中存在的问题中的至少一个,更好的满足内容匹配的需求,本申请提供了一种内容匹配方法,基于该方法可以提升内容匹配的准确率。

[0063] 为使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请实施方式作进一步地详细描述。

[0064] 本申请实施例是内容匹配系统提供的一种内容匹配方法,该内容匹配方法涉及人工智能领域,例如机器学习、深度学习等技术。示例性的,本申请实施例中所涉及的模型可以利用机器学习包括的人工神经网络技术;又例如,本申请实施例中所涉及的内容匹配可以利用深度学习以实现内容中文本和图像的识别。

[0065] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0066] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习、自动驾驶、智慧交通等几大方向。

[0067] 机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、

逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0068] 深度学习(Deep Learning, DL)是机器学习领域中一个新的研究方向。深度学习是学习样本数据的内在规律和表示层次,这些学习过程中获得的信息对诸如文字、图像和声音等数据的解释有很大的帮助。深度学习的最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据。

[0069] 为了更好的理解及说明本申请实施例的方案,下面对本申请实施例中所涉及到的一些技术用语进行简单说明。

[0070] TRANSFORMER: TRANSFORMER最初是作为机器翻译的Seq2Seq模型提出的,基于TRANSFORMER的预训练模型已成为NLP(Natural Language Processing,自然语言处理)中的首选架构;各种TRANSFORMER变体(X-TRANSFORMER)从不同的角度改进了原版TRANSFORMER。

[0071] 维基百科: 维基百科是一个基于维基技术的多语言百科全书式的协作计划,是用多种语言编写而成的网络百科全书。

[0072] XLNet: XLNet是一种通用的自回归预训练方法。

[0073] One-vs-All: One-vs-All是一对多算法,One-vs-All使用多个分类器,对于每一个类假设其为正类,然后使用一次二分类的判别算法即可分别得出各个类的分类。

[0074] Zero-shot learning: 在传统的分类模型中,为了解决多分类问题,例如三个类别:猫、狗和猪,就需要提供大量的猫、狗和猪的图像用以模型训练,然后给定一张新的图像,就能判定属于猫、狗或猪的其中哪一类。但是对于之前训练图像未出现的类别,例如牛,这个模型便无法将牛识别出来,而Zero-shot learning(零样本学习)就是为了解决这种问题。在Zero-shot learning中,某一类别在训练样本中未出现,但是知道这个类别的特征,然后通过语料知识库,便可以将这个类别识别出来。

[0075] CLIP模型: CLIP对比图文预训练(Contrastive Language-Image Pretraining,对比图文预训练)模型采用的是经典的双塔结构,对于图像域和文本域有着不同的图像编码器(Image Encoder)和文本编码器(Text Encoder);其中,文本编码器采用了经典的Transformer结构,而图像编码器则采用了两种:第一种图像编码器是改进后的残差网络ResNet,选择用基于注意力的池化层去替代ResNet的全局池化层;采用ViT结构作为第二种图像编码器进行实验。

[0076] Batch Size: 一次训练所选取的样本数。

[0077] NeXtVLAD: NeXtVLAD是将帧级别的特征降维成视频级别的特征,然后可以对视频级别的特征进行分类;NeXtVLAD先将高维的特征分解成一组低维的特征,然后加入注意力机制,再进行特征的编解码,最终达到降维的效果。

[0078] BERT模型: BERT(Bidirectional Encoder Representations from Transformers,基于转换器的双向编码表征)模型是一个语言表征模型,通过超大数据和极大的计算开销训练而成。

[0079] SE模块:SE(Squeeze and Excitation,压缩和激发)模块在于关注通道channel之间的关系,希望模型可以自动学习到不同channel特征的重要程度。

[0080] k-means聚类算法:k均值聚类算法(k-means clustering algorithm)是一种迭代求解的聚类分析算法,其步骤是,预将数据分为K组,则随机选取K个对象作为初始的聚类中心,然后计算每个对象与各个种子聚类中心之间的距离,把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本,聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有(或最小数目)对象被重新分配给不同的聚类,没有(或最小数目)聚类中心再发生变化,误差平方和局部最小。

[0081] 本申请实施例提供的方案涉及人工智能,下面以具体的实施例对本申请的技术方案以及本申请的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图,对本申请的实施例进行描述。

[0082] 本申请实施例所提供的方案可以适用于人工智能领域中任意需要内容匹配的应用场景,例如文章检索。

[0083] 为了更好的理解本申请实施例提供的方案,下面结合具体的一个应用场景对该方案进行说明。

[0084] 在一个实施例中,图1中示出了本申请实施例所适用的一种内容匹配系统的架构示意图,可以理解的是,本申请实施例所提供的内容匹配方法可以适用于但不限于应用于如图1所示的应用场景中。

[0085] 本示例中,如图1所示,该示例中的内容匹配系统的架构可以包括但不限于内容匹配平台10和数据库系统20,其中,内容匹配平台10可以为服务器或终端,数据库系统20可以为服务器;内容匹配平台10和数据库系统20之间可以通过网络进行交互。内容匹配平台10中运行神经网络模型110,神经网络模型110包括文本特征提取模型111、图像特征提取模型112、第一文本图像特征融合模型113、第二文本图像特征融合模型114和加权拼接模型115。

[0086] 将待查询内容的文本输入至文本特征提取模型111,确定待查询内容中文本的第一特征向量,将待查询内容的各图像分别输入至图像特征提取模型112,确定待查询内容中各图像的第二特征向量;将各图像的第二特征向量输入至第一文本图像特征融合模型113,将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量输入至第二文本图像特征融合模型114和加权拼接模型115,进行融合加权处理,确定待查询内容的特征嵌入向量;根据待查询内容的特征嵌入向量和数据库系统20中多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容;其中,内容可以为文章,图像可以为图片、视频帧等。

[0087] 可理解,上述仅为一种示例,本实施例在此不作限定。

[0088] 其中,终端可以是智能手机(如Android手机、iOS手机等)、手机模拟器、平板电脑、笔记本电脑、数字广播接收器、MID(Mobile Internet Devices,移动互联网设备)、PDA(个人数字助理)、车载终端(例如车载导航终端)、智能音箱、智能手表等。服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全

服务、CDN(Content Delivery Network,内容分发网络)、以及大数据和人工智能平台等基础云计算服务的云服务器或服务器集群。上述网络可以包括但不限于:有线网络,无线网络,其中,该有线网络包括:局域网、城域网和广域网,该无线网络包括:蓝牙、Wi-Fi及其他实现无线通信的网络。具体也可基于实际应用场景需求确定,在此不作限定。

[0089] 参见图2,图2示出了本申请实施例提供的一种内容匹配方法的流程示意图,其中,该方法可以由任一电子设备执行,如可以是服务器或终端,作为一可选实施方式,该方法可以由服务器或终端执行,为了描述方便,在下文的一些可选实施例的描述中,将以服务器或终端作为该方法执行主体为例进行说明。如图2所示,本申请实施例提供的内容匹配方法包括如下步骤:

[0090] S201,确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征。

[0091] 具体地,内容可以为文章,图像可以为图片、视频帧等。例如,待查询内容中文本可以为“西瓜是一种好吃的水果”、“花园里有只斑点狗”等,待查询内容中各图像可以为西瓜、人、狗、猫、鸟等图像。又例如,一个待查询内容中文本为“西瓜是一种好吃的水果”,该待查询内容中各图像为西瓜图像。第一特征向量为文本的向量化细粒度表征,其中,文本的向量化细粒度表征可以用于表示基于聚类的词的粒度,例如,文本的第一特征向量的尺寸大小为 1×128 ,即文本的向量化细粒度表征的尺寸大小为 1×128 。

[0092] S202,将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量。

[0093] 具体地,将多个图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量,其中,待查询内容的图像特征向量可以为embedding向量。

[0094] S203,将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量。

[0095] 具体地,待查询内容的特征嵌入向量用于表征待查询内容的特征嵌入embedding。融合加权处理包括融合处理和加权处理;将图像特征向量和第一特征向量之间进行融合处理,得到待查询内容的N维初始特征嵌入向量;将N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待查询内容的特征嵌入向量;其中,N为正整数。

[0096] S204,根据待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容。

[0097] 具体地,针对该待查询内容进行内容匹配,得到数据库中与待查询内容对应的相似内容,待查询内容与相似内容属于同一类内容。例如,一个待查询内容中文本为“西瓜是一种好吃的水果”,该待查询内容中各图像为西瓜图像;相似内容中文本为“西瓜是一种水果”,相似内容中各图像为西瓜图像。

[0098] 本申请实施例中,确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;如此,将文本的第一特征向量和图像特征向量之间进行融合加权处理,实现了待查询内容的特征嵌

入向量的提取,即待查询内容的多模态特征的提取,从而通过待查询内容的特征嵌入向量,更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

[0099] 在一个实施例中,确定待查询内容中文本的第一特征向量,包括步骤A1-A2:

[0100] 步骤A1,将待查询内容中文本输入至神经网络模型的文本特征提取模型的匹配子模型,将文本和预设的多个聚类集合经过匹配处理,确定文本的第一特征嵌入向量对应的一个聚类集合,多个聚类集合包括一个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签。

[0101] 在一个实施例中,神经网络模型的文本特征提取模型可以为基于监督训练的文本特征提取模型XBERT, XBERT对BERT模型无法满足文本细粒度特征的缺点进行了改进,加入了X-TRANSFORMER模块, X-TRANSFORMER模块可以使XBERT能够关注细粒度的文本标签。

[0102] 在一个实施例中, XBERT可以包括语义标签索引模块、深度神经匹配模块和集成排序模块;其中,语义标签索引模块用于建立包括多个聚类集合的索引系统,聚类集合可以作为索引系统中的索引;深度神经匹配模块可以为神经网络模型的文本特征提取模型XBERT的匹配子模型,深度神经匹配模块将文本和预设的多个聚类集合经过匹配处理,确定文本的第一特征嵌入向量对应的一个聚类集合。

[0103] 在一个实施例中,深度神经匹配模块是将相关的聚类集合分配给每个实例,其中,实例可以是内容的文本实例。在多个文本标签聚类之后,多个文本标签被划分为K个聚类集合。深度神经匹配模块的目标是找到一个编码器g来创建一个实例的特征嵌入向量embedding $u=g(x)$,并学习将该实例的embedding u 映射到相关聚类集合的浅层神经网络。例如,深度神经匹配模块可以是深度TRANSFORMER模型,深度神经匹配模块将每个实例的特征映射到一个聚类集合,可以用实例-聚类分配矩阵表示,实例-聚类分配矩阵用于表示实例和聚类集合之间的相关度,如公式(1)所示。

[0104] $M=YC=[m_1, \dots, m_i, \dots, m_N]^T \in \{0, 1\}^{N \times K}$ 公式(1)

[0105] 其中,M是实例-聚类分配矩阵,Y是实例-文本标签分配矩阵,C是文本标签-聚类矩阵; $m_1, \dots, m_i, \dots, m_N$ 表示实例和聚类集合之间的相关度,N表示N个实例,K表示K个聚类集合, $\{0, 1\}$ 表示实例和聚类集合之间相关度的取值范围,N和K都为正整数。

[0106] 步骤A2,根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量。

[0107] 在一个实施例中, XBERT包括的集成排序模块根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量。

[0108] 在一个实施例中,根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量,包括步骤B1-B2:

[0109] 步骤B1,确定文本的第一特征嵌入向量和各文本标签的特征嵌入向量之间的第一相似度。

[0110] 在一个实施例中,计算文本的第一特征嵌入向量分别和每个文本标签的特征嵌入向量之间的第一相似度,得到多个第一相似度,其中,每个文本标签都归属于同一个聚类集合。

[0111] 步骤B2,根据各文本标签的特征嵌入向量和各第一相似度对应的权重,确定第一特征向量,各第一相似度与各第一相似度对应的权重之间呈正相关。

[0112] 在一个实施例中,第一相似度越大,则该第一相似度对应的权重也越大。计算各文本标签的特征嵌入向量和各第一相似度对应的权重之间的乘积,得到各文本标签对应的向量加权值;计算各文本标签对应的向量加权值之间的和,得到第一特征向量。

[0113] 在一个实施例中,根据各文本标签的特征嵌入向量和各第一相似度对应的权重,确定第一特征向量,包括:

[0114] 将各第一相似度输入至神经网络模型的文本特征提取模型的排序子模型,将各第一相似度从大到小进行排序,确定排序在前的M个第一相似度;

[0115] 根据M个第一相似度对应的权重,以及M个第一相似度对应的文本标签的特征嵌入向量,确定第一特征向量,M为正整数。

[0116] 在一个实施例中,计算M个第一相似度对应的文本标签中各文本标签的特征嵌入向量和各第一相似度对应的权重之间的乘积,得到M个第一相似度对应的文本标签中各文本标签对应的向量加权值;计算M个向量加权值之间的和,得到第一特征向量。

[0117] 在一个实施例中,XBERT包括的集成排序模块对聚类集合中的文本标签进行排序。例如,给定一个文本标签1和一个实例x,集成排序模块找到一个映射 $h(x, 1)$,将实例x和文本标签1映射到一个分数,主要使用One-vs-All的方法,One-vs-All将单个文本标签对实例的分配视为一个独立的二分类问题;当实例x归属于聚类集合,则文本标签1为正,否则文本标签1为负。

[0118] 可以进一步对不同类型的XBERT模型的得分进行集成,例如,针对多个不同类型的XBERT模型,标签1可能分别得到不同的分数,则将各分数进行加权,得到标签1的最后得分。

[0119] 在一个实施例中,在确定待查询内容中文本的第一特征向量之前,还包括步骤C1-C2:

[0120] 步骤C1,获取多个文本标签。

[0121] 在一个实施例中,XBERT包括的语义标签索引模块可以找到有意义的文本标签表示,以建立索引系统。生成文本标签特征嵌入向量embedding(文本标签embedding)包括以下两种方式:

[0122] 方式1,通过标签文本生成文本标签embedding。

[0123] 具体地,给定有关标签的文本信息,例如维基百科数据集中类别的简短描述,可以使用此短文本来表示标签词。可以使用XLNet模型来生成标签词1,标签词1的文本标签embedding可以通过在标签文本中的所有词向量平均池化创建的,如公式(2)所示。

[0124]
$$\Psi_{\text{text-emb}}(l) = \frac{1}{|\text{text}(l)|} \sum_{w \in \text{text}(l)} \phi_{\text{xlnet}}(w)$$
 公式(2)

[0125] 其中, $\Psi_{\text{text-emb}}(l)$ 表示标签词1的文本标签embedding, $\text{text}(l)$ 表示w的集合,w表示对标签词1进行解释的文本(例如,标签词1是西瓜,w是“一个好吃的水果”), $|\text{text}(l)|$ 表示w的数量,l表示标签词1, $\phi_{\text{xlnet}}(w)$ 表示一个w的特征嵌入向量embedding。

[0126] 方式2,通过正例(正样本)中的关键字生成文本标签embedding。

[0127] 具体地,标签的短文本信息可能不包含足够的信息,并且短文本中的某些单词可能有歧义和噪声。因此,考虑从实例的稀疏文本embedding导出的另一个标签表示,例如,文本标签embedding是标签词1(标签词1例如西瓜,一个短文本是“一个好吃的水果”,其中,“好吃的”对应的特征是稀疏TF-IDF特征)的所有内容的相关实例的稀疏TF-IDF特征的总

和,如公式(3)所示。

$$[0128] \quad \Psi_{\text{pifa-tfidf}}(l) = \frac{v_l}{\|v_l\|}, v_l = \sum_{i: y_{il}=1} \phi_{\text{tf-idf}}(x_i), l=1, \dots, L \text{ 公式(3)}$$

[0129] 其中, $\Psi_{\text{pifa-tfidf}}(l)$ 表示文本标签embedding; $\|v_l\|$ 表示多个实例的数量, y_{il} 表示所有标签词的总数, $\phi_{\text{tf-idf}}(x_i)$ 表示一个实例 x_i 的稀疏TF-IDF特征。

[0130] 步骤C2,将多个文本标签进行聚类处理,得到多个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签。

[0131] 在一个实施例中,通过k-means聚类算法,对多个文本标签embedding进行聚类来建立索引系统,索引系统包括多个聚类集合。

[0132] 在一个实施例中,确定待查询内容中各图像的第二特征向量,包括:

[0133] 将待查询内容中各图像分别输入至神经网络模型的图像特征提取模型的训练后的图像编码器,对各图像分别进行编码处理,得到待查询内容中各图像的第二特征向量。

[0134] 在一个实施例中,神经网络模型的图像特征提取模型可以为基于自监督训练的对比图文预训练CLIP模型。CLIP模型的推理阶段(应用阶段)包括:将待查询内容中各图像分别输入至训练后的图像编码器,对各图像分别进行编码处理,得到待查询内容中各图像的第二特征向量。例如,将一张图像通过训练后的CLIP模型的图像编码器,提取该图像的尺寸大小为 1×512 的图像特征,即该图像的第二特征向量。

[0135] 在一个实施例中,在确定待查询内容中各图像的第二特征向量之前,还包括:

[0136] 获取预设数据集中多个文本样本和多个图像样本;

[0137] 将多个文本样本输入至所述图像特征提取模型的文本编码器,并将多个图像样本输入至图像特征提取模型的图像编码器,进行对比学习处理,得到多个正样本对,多个正样本对中的每个正样本对包括一个图像样本和一个文本样本,一个图像样本和一个文本样本之间的第二相似度大于或等于预设相似度阈值;

[0138] 确定各正样本对的损失函数值,损失函数值用于表征各正样本对中图像样本的特征嵌入向量和文本样本的特征嵌入向量之间的距离;

[0139] 基于各正样本对的损失函数值,更新文本编码器和所述图像编码器;

[0140] 若各正样本对的损失函数值小于或等于预设损失函数值阈值,则结束文本编码器和图像编码器的训练,并将训练得到的图像编码器作为图像特征提取模型的训练后的图像编码器。

[0141] 在一个实施例中,神经网络模型的图像特征提取模型可以为基于自监督训练的对比图文预训练CLIP模型。CLIP模型通过对比损失,配对图像embedding与文本embedding,来训练图像编码器和文本编码器,其中,图像编码器使用ResNet或ViT。

[0142] 例如,CLIP模型的训练数据集包括:4亿对构造,50万词条(类别);batch size $N=32768$,32768是2的15次方,每个batch,有 N^2 图像-文本对,其中 N 个图像-文本对(例如,一个图像-文本对,其中,图像是西瓜的图像,文本是西瓜)是正样本,希望正样本的图像embedding和文本embedding之间距离越近越好,其余的 N^2-N 个是负样本,希望可以最小化其embedding相似度;其中,图像-文本对可以为图片-文本对。

[0143] 例如,CLIP模型的训练阶段包括:通过预训练图像编码器和文本编码器,以对比学习的方式预测数据集中哪些图像与哪些文本配对,计算对比损失,例如交叉熵(cross

entropy) 损失,交叉熵损失越小,图像与文本越匹配。

[0144] 需要说明的是,CLIP模型的工作流程包括:通过预训练图像编码器和文本编码器,以对比学习的方式预测数据集中哪些图像与哪些文本配对,然后,将CLIP模型转换为Zero-shot learning分类器。利用自监督训练的CLIP图像特征(各图像的第二特征向量)作为对标签监督训练的文本特征(文本的第一特征向量)的一个补充,以使得后续神经网络模型提取到待查询内容的特征嵌入向量,一方面获得扩展性更好的特征信息,另一方面补足图像特征缺失的不足,实现了待查询内容的多模态特征的提取,从而更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

[0145] 在一个实施例中,将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量,包括步骤:

[0146] 将各图像的第二特征向量输入至神经网络模型的第一文本图像特征融合模型,进行融合处理,得到待查询内容的图像特征向量;

[0147] 神经网络模型还包括文本特征提取模型、图像特征提取模型、第二文本图像特征融合模型和加权拼接模型。

[0148] 在一个实施例中,神经网络模型的第一文本图像特征融合模型可以为聚合网络,例如NeXtVLAD。

[0149] 举例说明,待查询内容包括5个图像,这5个图像依次输入CLIP模型分别进行处理,得到5个图像特征,即5个图像的第二特征向量,这5个图像的第二特征向量的尺寸大小都为 1×512 ;将这5个图像的第二特征向量通过NeXtVLAD,进行融合处理,聚合成一个内容级别的图像特征,即待查询内容的图像特征向量,待查询内容的图像特征向量的尺寸大小为 1×512 。

[0150] 一个实施例中,将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量,包括步骤D1-D2:

[0151] 步骤D1,将图像特征向量和第一特征向量输入至神经网络模型的第二文本图像特征融合模型,进行融合处理,得到待查询内容的N维初始特征嵌入向量。

[0152] 在一个实施例中,神经网络模型的第二文本图像特征融合模型可以为聚合网络,例如NeXtVLAD。

[0153] 举例说明,待查询内容的文本输入XBERT,得到一个内容级别的文本特征嵌入向量,即文本的第一特征向量,文本的第一特征向量的尺寸大小为 1×128 ;待查询内容的图像特征向量的尺寸大小为 1×512 ;将文本的第一特征向量和待查询内容的图像特征向量通过NeXtVLAD,进行融合处理,聚合成一个内容级别的特征嵌入向量,即待查询内容的128维初始特征嵌入向量,待查询内容的128维初始特征嵌入向量的尺寸大小为 1×128 ,N取值为128。

[0154] 步骤D2,将待查询内容的N维初始特征嵌入向量输入至神经网络模型的加权拼接模型,将N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待查询内容的特征嵌入向量,N为正整数。

[0155] 在一个实施例中,神经网络模型的加权拼接模型可以为SE模块。

[0156] 举例说明,将待查询内容的128维初始特征嵌入向量输入至SE模块,将128维初始特征嵌入向量对应的128个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待

查询内容的特征嵌入向量,待查询内容的特征嵌入向量可以表示待查询内容的多模态特征向量。

[0157] 应用本申请实施例,至少具有如下有益效果:

[0158] 将基于监督训练的神经网络模型的文本特征提取模型和基于自监督训练的神经网络模型的图像特征提取模型相结合,将两者提取的特征(第一特征向量和图像特征向量)统一成结合文本和视觉的特征表达(待查询内容的特征嵌入向量)。一方面,利用文本特征提取模型提取标签相关的细粒度表观特征,另一方面,利用图像特征提取模型提取信息更全面和扩展性更好的图像特征,同时能避免监督训练中由于监督信号本身的噪声和不合理性所带来的问题,从而更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

[0159] 为了更好的理解本申请实施例所提供的方法,下面结合具体应用场景的示例对本申请实施例的方案进行进一步说明。

[0160] 内容匹配的应用场景可以是文章检索,本申请实施例所提供的文章检索应用于审核系统中的图文特征提取,例如进行图文文章排重召回、维权文章相似召回、原创文章相似召回等。如图3所示,本申请实施例提供的原创文章相似召回,包括如下步骤:

[0161] S301,获取声明的原创文章。

[0162] S302,将声明的原创文章输入至神经网络模型,确定声明的原创文章的特征嵌入向量。

[0163] 在一个实施例中,如图1所示,将声明的原创文章的文本输入至文本特征提取模型111,确定声明的原创文章中文本的第一特征向量,将声明的原创文章的各图像分别输入至图像特征提取模型112,确定声明的原创文章中各图像的第二特征向量;将各图像的第二特征向量输入至第一文本图像特征融合模型113,将各图像的第二特征向量之间进行融合处理,确定声明的原创文章的图像特征向量;将图像特征向量和第一特征向量输入至第二文本图像特征融合模型114和加权拼接模型115,进行融合加权处理,确定声明的原创文章的特征嵌入向量。

[0164] 举例说明,神经网络模型110的文本特征提取模型111可以为基于监督训练的文本特征提取模型XBERT;神经网络模型110的图像特征提取模型112可以为基于自监督训练的对比图文预训练CLIP模型;神经网络模型110的第一文本图像特征融合模型113可以为聚合网络,例如NeXtVLAD;神经网络模型110的第二文本图像特征融合模型114可以为聚合网络,例如NeXtVLAD;神经网络模型110的加权拼接模型115可以为SE模块。

[0165] S303,根据声明的原创文章的特征嵌入向量和预设数据库中多个文章的特征嵌入向量,判断声明的原创文章和多个文章之间是否有相似内容;当确定声明的原创文章和多个文章之间没有相似内容,则转到步骤S304处理;当确定声明的原创文章和多个文章之间有相似内容,则转到步骤S305处理。

[0166] S304,针对声明的原创文章的相似审核通过,确定声明的原创文章为原创文章。

[0167] S305,判断相似内容的发文时间是否早于声明的原创文章的发文时间,当确定相似内容的发文时间晚于声明的原创文章的发文时间,则转到步骤S304处理;当确定相似内容的发文时间是早于声明的原创文章的发文时间,则转到步骤S306处理。

[0168] S306,判断发文时间早于声明的原创文章的相似内容的作者是否与声明的原创文

章的作者相同;当确定两者的作者相同,则转到步骤S307处理;当确定两者的作者不相同,则转到步骤S308处理。

[0169] S307,判断发文时间早于声明的原创文章的相似内容是否有原创标记,当确定发文时间早于声明的原创文章的相似内容没有原创标记,则转到步骤S304处理;当确定发文时间早于声明的原创文章的相似内容有原创标记,则转到步骤S308处理。

[0170] S308,针对声明的原创文章的相似审核不通过,确定声明的原创文章不为原创文章。

[0171] 在一个实施例中,在推荐场景中,本申请实施例所提供的内容匹配方法可以帮助业务方将图文标签应用在推荐系统的各个环节,例如丰富用户长短期兴趣、对文章进行召回排序、提高冷启动分发能力等。

[0172] 在一个实施例中,例如,图文标签模型的基线指标为准确率78%,召回率57%。图文标签模型加上XBERT细粒度的文本特征后,准确率提升至80%,召回率提升至58%;在此基础上融合CLIP提取的图像特征,准确率提升至83%,召回率提升至60%。不同模型的效果对比如表1所示,图文标签模型在加入XBERT和CLIP之后,相对于之前图文标签模型的准确率和召回率有了明显的提高。

[0173] 表1不同模型的效果对比

模型	准确率	召回率
图文标签模型	78%	57%
图文标签模型+XBERT	80%	58%
图文标签模型+XBERT+CLIP	83%	60%

[0175] 参见图4,图4示出了本申请实施例提供的一种文章检索的示意图,其中,该方法可以由任一电子设备执行,如可以是服务器或终端,作为一可选实施方式,该方法可以由服务器或终端执行,为了描述方便,在下文的一些可选实施例的描述中,将以服务器或终端作为该方法执行主体为例进行说明。如图4所示,本申请实施例提供的文章检索包括如下步骤:

[0176] S401,获取待查询文章。

[0177] 需要说明的是,待查询文章为待查询内容。

[0178] S402,将待查询文章的文本输入至基于监督训练的文本特征提取模型XBERT,确定待查询文章中文本的第一特征向量。

[0179] 需要说明的是,XBERT对BERT模型无法满足文本细粒度特征的缺点进行了改进,加入了X-TRANSFORMER模块,X-TRANSFORMER模块可以使XBERT能够关注细粒度的文本标签。

[0180] S403,将待查询文章各图像分别输入至基于自监督训练的对比图文预训练CLIP模型,确定待查询文章中各图像的第二特征向量。

[0181] S404,将各图像的第二特征向量输入至第一NeXtVLAD,将各图像的第二特征向量之间进行融合处理,确定待查询文章的图像特征向量。

[0182] S405,将图像特征向量和第一特征向量输入至第二NeXtVLAD,进行融合处理,得到待查询文章的N维初始特征嵌入向量,其中,N为正整数。

[0183] S406,将待查询文章的N维初始特征嵌入向量输入至SE模块,将N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待查询文章的特征嵌入向量。

[0184] S407,根据待查询文章的特征嵌入向量和数据库系统中多个文章的特征嵌入向量,判断待查询文章和多个文章之间是否有相似内容。

[0185] 应用本申请实施例,至少具有如下有益效果:

[0186] 使用人工标签作为监督信号训练XBERT,并在此基础上加入X-TRANSFORMER模块使其能提取具有丰富语义的细粒度标签特征。利用图像自监督模型CLIP提取图像特征,以获得信息全面和扩展性较好的视觉特征表达,同时减少对人工标注的依赖。结合了以人工标签作为监督信号来训练的模型XBERT提取的文本特征和以自监督方式训练的图像模型CLIP提取的视觉特征,两者作为多模态特征进行训练融合,可以供下一阶段模型或下游任务使用。实现了待查询文章的特征嵌入向量的提取,即待查询文章的多模态特征的提取,从而通过待查询文章的特征嵌入向量,更全面的对文章的文本和图像进行综合表征,提升了文章检索的准确率。

[0187] 本申请实施例还提供了一种内容匹配装置,该内容匹配装置的结构示意图如图5所示,内容匹配装置60,包括第一处理模块601、第二处理模块602、第三处理模块603和第四处理模块604。

[0188] 第一处理模块601,用于确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;

[0189] 第二处理模块602,用于将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;

[0190] 第三处理模块603,用于将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;

[0191] 第四处理模块604,用于根据待查询内容的特征嵌入向量和预设的多个内容的特征嵌入向量,从多个内容中匹配待查询内容对应的相似内容。

[0192] 在一个实施例中,第一处理模块601,具体用于:

[0193] 将待查询内容中文本输入至神经网络模型的文本特征提取模型的匹配子模型,将文本和预设的多个聚类集合经过匹配处理,确定文本的第一特征嵌入向量对应的一个聚类集合,多个聚类集合包括一个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签;

[0194] 根据文本的第一特征嵌入向量和聚类集合中各文本标签的特征嵌入向量,确定第一特征向量。

[0195] 在一个实施例中,第一处理模块601,具体用于:

[0196] 确定文本的第一特征嵌入向量和各文本标签的特征嵌入向量之间的第一相似度;

[0197] 根据各文本标签的特征嵌入向量和各第一相似度对应的权重,确定第一特征向量,各第一相似度与各第一相似度对应的权重之间呈正相关。

[0198] 在一个实施例中,第一处理模块601,具体用于:

[0199] 将各第一相似度输入至神经网络模型的文本特征提取模型的排序子模型,将各第

一相似度从大到小进行排序,确定排序在前的M个第一相似度;

[0200] 根据M个第一相似度对应的权重,以及M个第一相似度对应的文本标签的特征嵌入向量,确定第一特征向量,M为正整数。

[0201] 在一个实施例中,第一处理模块601,还用于:

[0202] 获取多个文本标签;

[0203] 将多个文本标签进行聚类处理,得到多个聚类集合,多个聚类集合中的每个聚类集合包括多个归属于同一类型的文本标签。

[0204] 在一个实施例中,第一处理模块601,具体用于:

[0205] 将待查询内容中各图像分别输入至神经网络模型的图像特征提取模型的训练后的图像编码器,对各图像分别进行编码处理,得到待查询内容中各图像的第二特征向量。

[0206] 在一个实施例中,第一处理模块601,还用于:

[0207] 获取预设数据集中多个文本样本和多个图像样本;

[0208] 将多个文本样本输入至所述图像特征提取模型的文本编码器,并将多个图像样本输入至图像特征提取模型的图像编码器,进行对比学习处理,得到多个正样本对,多个正样本对中的每个正样本对包括一个图像样本和一个文本样本,一个图像样本和一个文本样本之间的第二相似度大于或等于预设相似度阈值;

[0209] 确定各正样本对的损失函数值,损失函数值用于表征各正样本对中图像样本的特征嵌入向量和文本样本的特征嵌入向量之间的距离;

[0210] 基于各正样本对的损失函数值,更新文本编码器和所述图像编码器;

[0211] 若各正样本对的损失函数值小于或等于预设损失函数值阈值,则结束文本编码器和图像编码器的训练,并将训练得到的图像编码器作为图像特征提取模型的训练后的图像编码器。

[0212] 在一个实施例中,第二处理模块602,具体用于:

[0213] 将各图像的第二特征向量输入至神经网络模型的第一文本图像特征融合模型,进行融合处理,得到待查询内容的图像特征向量;

[0214] 神经网络模型还包括文本特征提取模型、图像特征提取模型、第二文本图像特征融合模型和加权拼接模型。

[0215] 在一个实施例中,第三处理模块603,具体用于:

[0216] 将图像特征向量和第一特征向量输入至神经网络模型的第二文本图像特征融合模型,进行融合处理,得到待查询内容的N维初始特征嵌入向量;

[0217] 将待查询内容的N维初始特征嵌入向量输入至神经网络模型的加权拼接模型,将N维初始特征嵌入向量对应的N个特征分别进行加权处理,并将加权后的各特征进行拼接,得到待查询内容的特征嵌入向量,N为正整数。

[0218] 应用本申请实施例,至少具有如下有益效果:

[0219] 确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;如此,将文本的第一特征向量和图像特征向量之间进行融合加权处理,实现了待查询内容的特征嵌入向量的提取,

即待查询内容的多模态特征的提取,从而通过待查询内容的特征嵌入向量,更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

[0220] 本申请实施例还提供了一种电子设备,该电子设备的结构示意图如图6所示,图6所示的电子设备4000包括:处理器4001和存储器4003。其中,处理器4001和存储器4003相连,如通过总线4002相连。可选地,电子设备4000还可以包括收发器4004,收发器4004可以用于该电子设备与其他电子设备之间的数据交互,如数据的发送和/或数据的接收等。需要说明的是,实际应用中收发器4004不限于一个,该电子设备4000的结构并不构成对本申请实施例的限定。

[0221] 处理器4001可以是CPU(Central Processing Unit,中央处理器),通用处理器,DSP(Digital Signal Processor,数据信号处理器),ASIC(Application Specific Integrated Circuit,专用集成电路),FPGA(Field Programmable Gate Array,现场可编程门阵列)或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框,模块和电路。处理器4001也可以是实现计算功能的组合,例如包含一个或多个微处理器组合,DSP和微处理器的组合等。

[0222] 总线4002可包括一通路,在上述组件之间传送信息。总线4002可以是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4002可以分为地址总线、数据总线、控制总线等。为便于表示,图6中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0223] 存储器4003可以是ROM(Read Only Memory,只读存储器)或可存储静态信息和指令的其他类型的静态存储设备,RAM(Random Access Memory,随机存取存储器)或者可存储信息和指令的其他类型的动态存储设备,也可以是EEPROM(Electrically Erasable Programmable Read Only Memory,电可擦可编程只读存储器)、CD-ROM(Compact Disc Read Only Memory,只读光盘)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质、其他磁存储设备、或者能够用于携带或存储计算机程序并能够由计算机读取的任何其他介质,在此不做限定。

[0224] 存储器4003用于存储执行本申请实施例的计算机程序,并由处理器4001来控制执行。处理器4001用于执行存储器4003中存储的计算机程序,以实现前述方法实施例所示的步骤。

[0225] 其中,电子设备包括但不限于:服务器、终端等。

[0226] 应用本申请实施例,至少具有如下有益效果:

[0227] 确定待查询内容中文本的第一特征向量和各图像的第二特征向量,第一特征向量为文本的向量化细粒度表征,第二特征向量用于表征各图像的视觉特征;将各图像的第二特征向量之间进行融合处理,确定待查询内容的图像特征向量;将图像特征向量和第一特征向量之间进行融合加权处理,确定待查询内容的特征嵌入向量;如此,将文本的第一特征向量和图像特征向量之间进行融合加权处理,实现了待查询内容的特征嵌入向量的提取,即待查询内容的多模态特征的提取,从而通过待查询内容的特征嵌入向量,更全面的对内容的文本和图像进行综合表征,提升了内容匹配的准确率。

[0228] 本申请实施例提供了一种计算机可读存储介质,该计算机可读存储介质上存储有计算机程序,计算机程序被处理器执行时可实现前述方法实施例的步骤及相应内容。

[0229] 本申请实施例还提供了一种计算机程序产品,包括计算机程序,计算机程序被处理器执行时可实现前述方法实施例的步骤及相应内容。

[0230] 基于与本申请实施例提供的方法相同的原理,本申请实施例还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述本申请任一可选实施例中提供的方法。

[0231] 应该理解的是,虽然本申请实施例的流程图中通过箭头指示各个操作步骤,但是这些步骤的实施顺序并不受限于箭头所指示的顺序。除非本文中有明确的说明,否则在本申请实施例的一些实施场景中,各流程图中的实施步骤可以按照需求以其他的顺序执行。此外,各流程图中的部分或全部步骤基于实际的实施场景,可以包括多个子步骤或者多个阶段。这些子步骤或者阶段中的部分或全部可以在同一时刻被执行,这些子步骤或者阶段中的每个子步骤或者阶段也可以分别在不同的时刻被执行。在执行时刻不同的场景下,这些子步骤或者阶段的执行顺序可以根据需求灵活配置,本申请实施例对此不限制。

[0232] 以上所述仅是本申请部分实施场景的可选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请的方案技术构思的前提下,采用基于本申请技术思想的其他类似实施手段,同样属于本申请实施例的保护范畴。

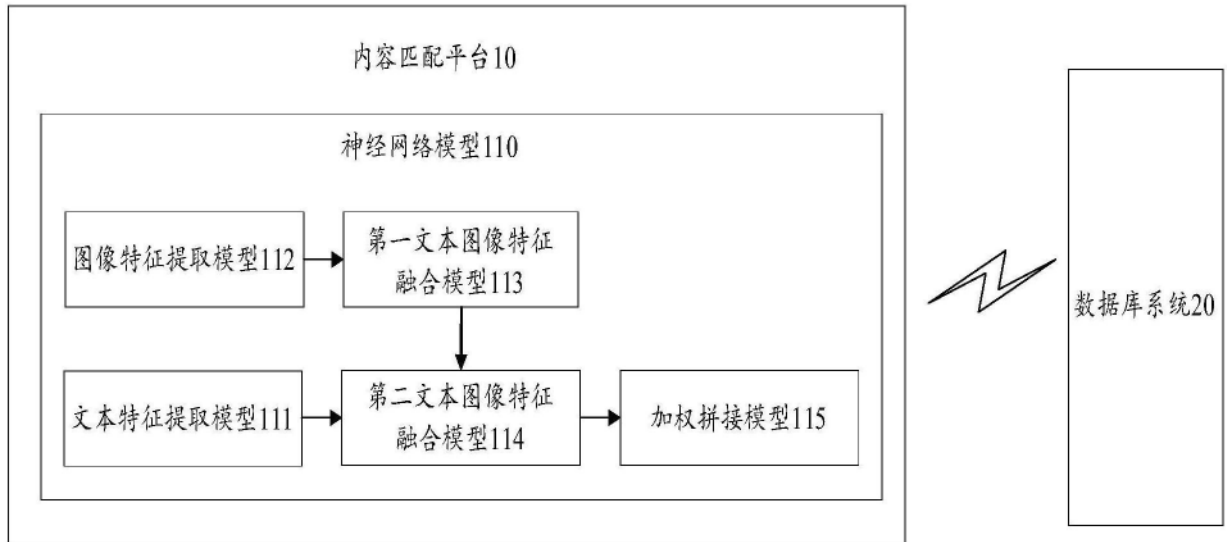


图1

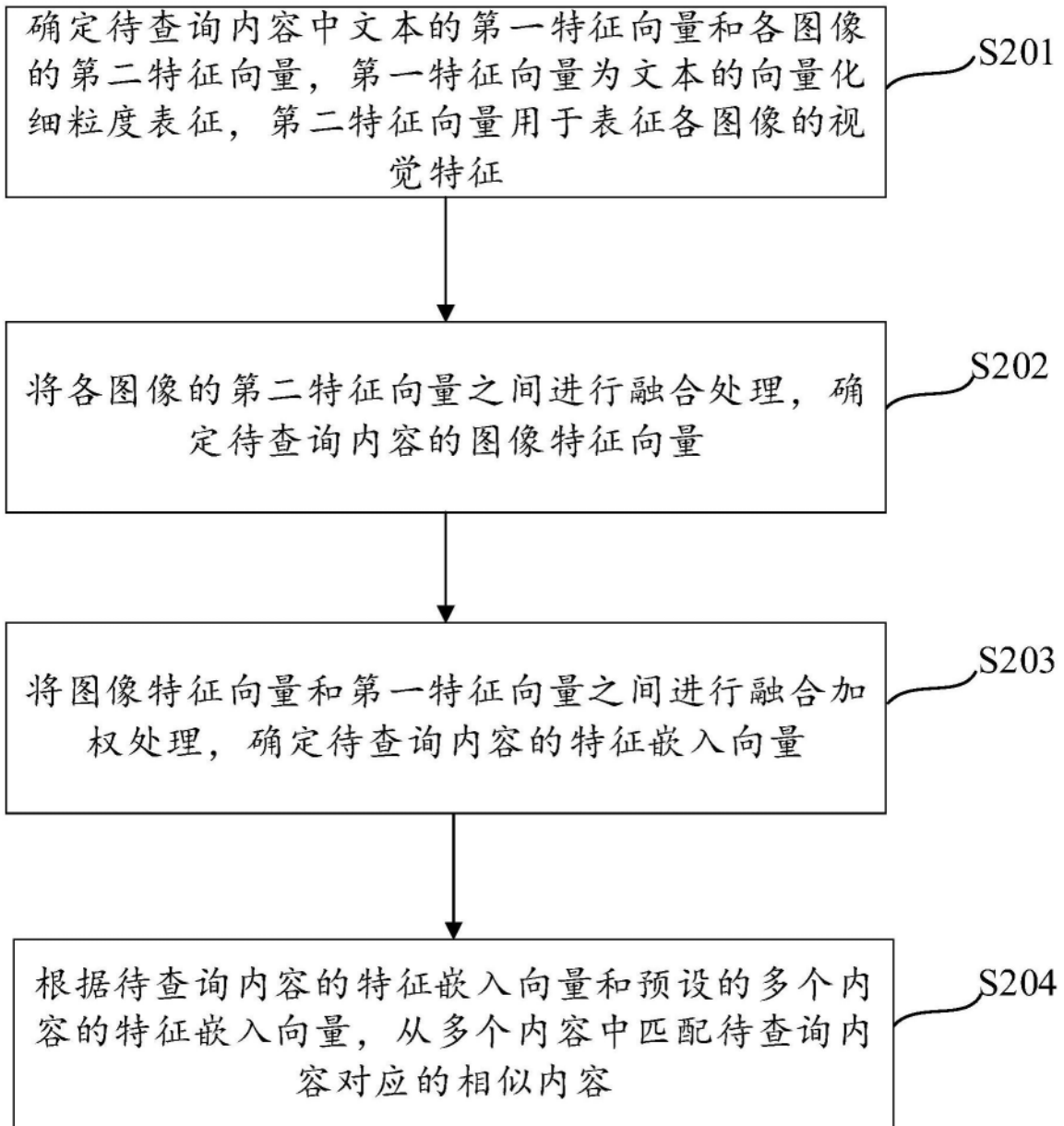


图2

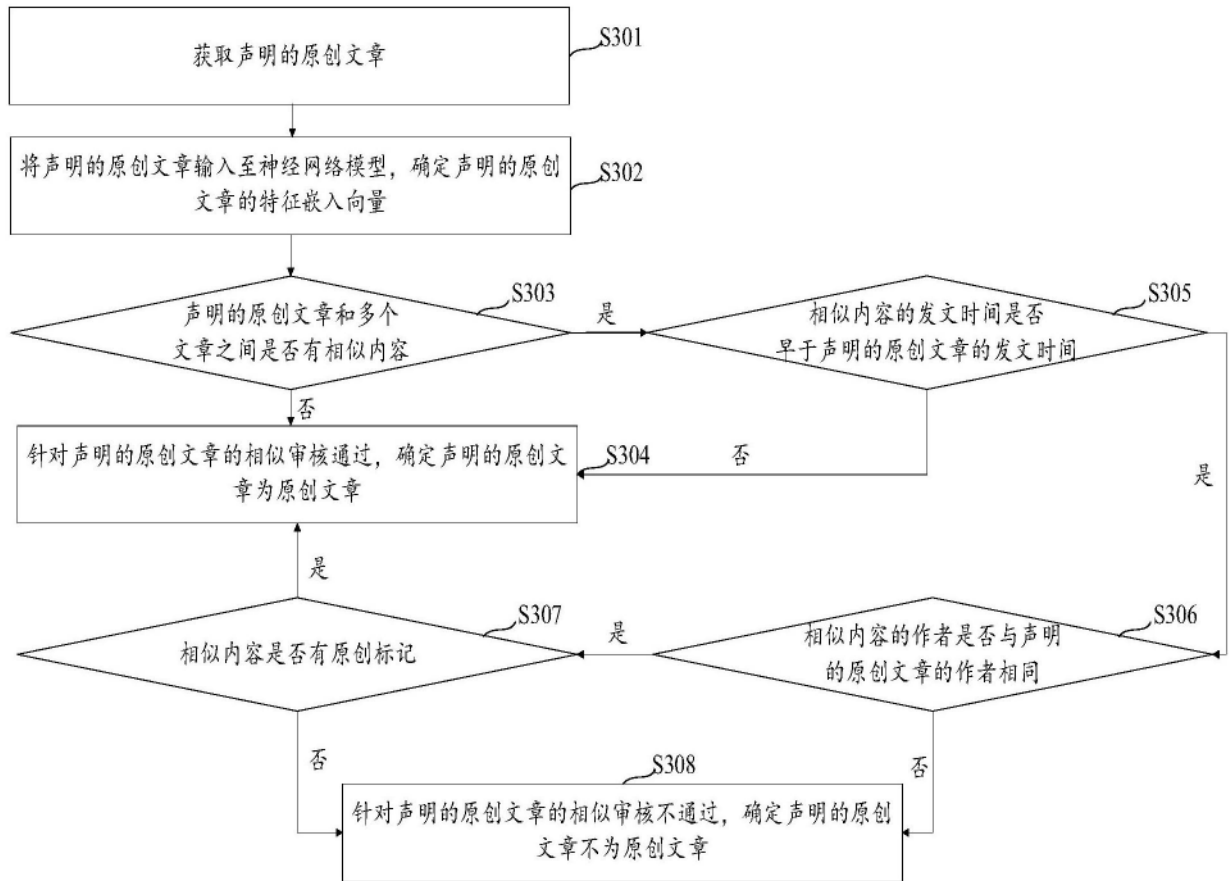


图3

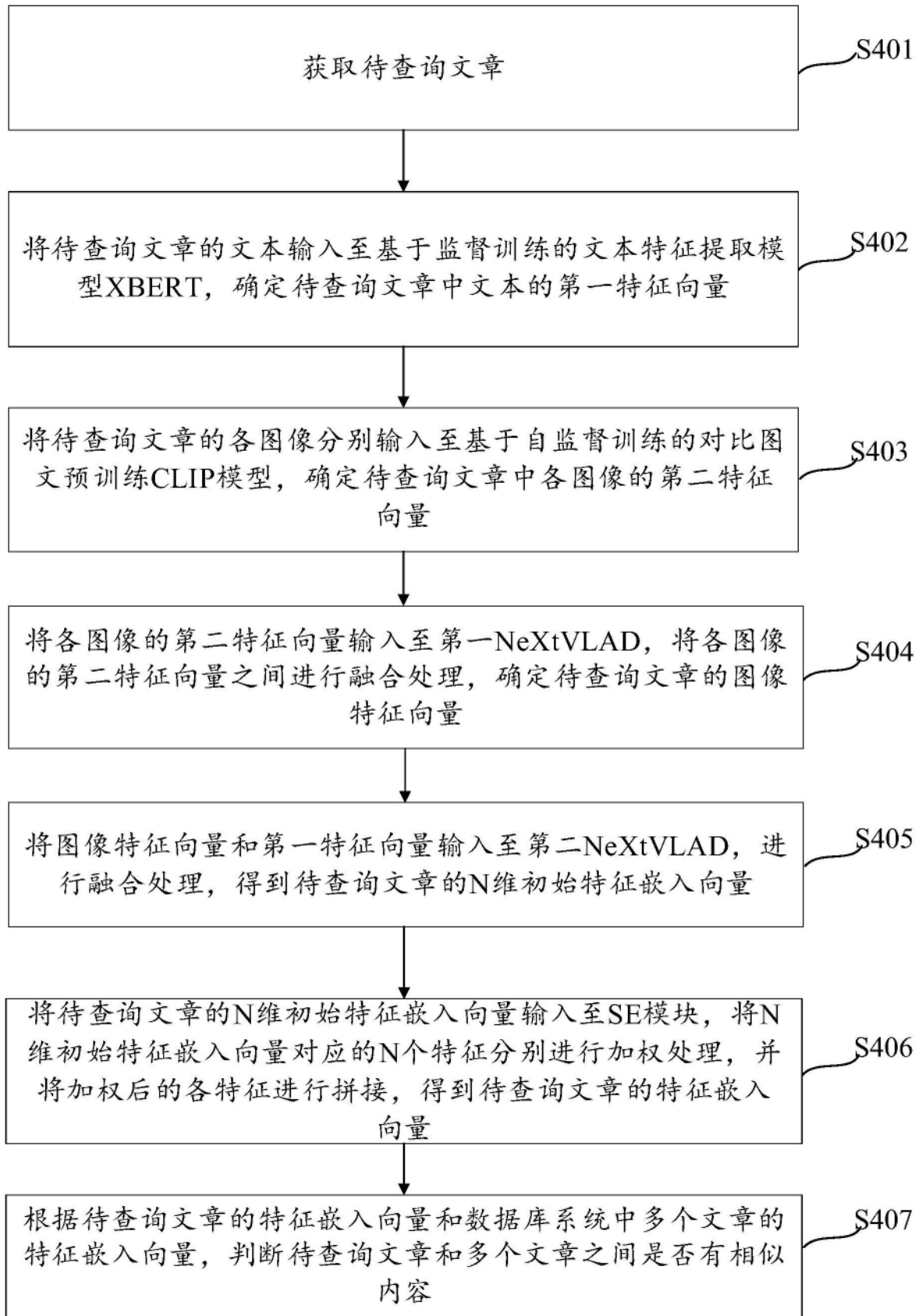


图4

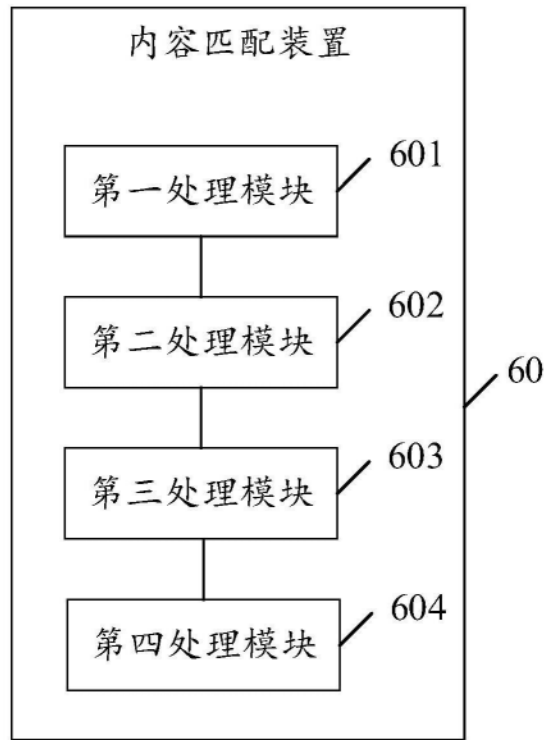


图5

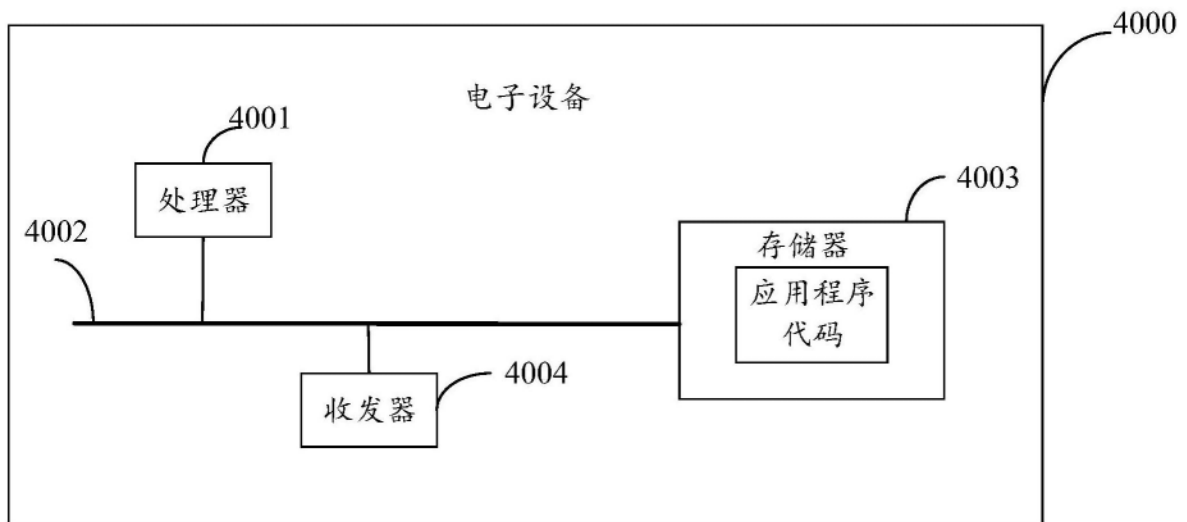


图6