



(12) 发明专利申请

(10) 申请公布号 CN 113010740 A

(43) 申请公布日 2021.06.22

(21) 申请号 202110258046.3

(22) 申请日 2021.03.09

(71) 申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72) 发明人 黄剑辉

(74) 专利代理机构 北京三高永信知识产权代理
有限责任公司 11138

代理人 祝亚男

(51) Int. Cl.

G06F 16/783 (2019.01)

G06F 40/258 (2020.01)

G06F 40/279 (2020.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

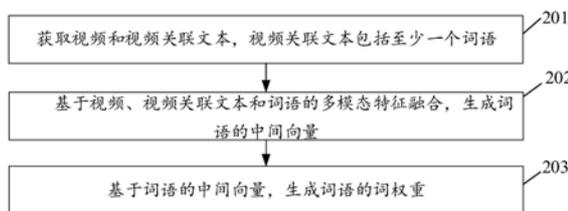
权利要求书2页 说明书17页 附图9页

(54) 发明名称

词权重的生成方法、装置、设备及介质

(57) 摘要

本申请公开了一种词权重的生成方法、装置、设备及介质,涉及信息处理领域。方法包括:获取视频和视频关联文本,视频关联文本包括至少一个词语;基于视频、视频关联文本和词语的多模态特征融合,生成词语的中间向量;基于词语的中间向量,生成词语的词权重。在视频搜索过程中,采用上述词权重生成方法来预先提取词语的权重值,不仅考虑了文本维度的特征,还引入融合了视频维度的特征,基于多维度的特征来进行词权重生成,有利于提升输出的词权重的准确率和可靠程度,提高了视频关联文本中对关键词语和混淆词语之间的区分度。



1. 一种词权重的生成方法,其特征在于,所述方法包括:
获取视频和视频关联文本,所述视频关联文本包括至少一个词语;
基于所述视频、所述视频关联文本和所述词语的多模态特征融合,生成所述词语的中间向量;
基于所述词语的中间向量,生成所述词语的词权重。
2. 根据权利要求1所述的方法,其特征在于,所述基于所述视频、所述视频关联文本和所述词语的多模态特征融合,生成所述词语的中间向量,包括:
提取所述视频的视频特征向量;提取所述视频关联文本的文本特征向量;以及提取所述词语的词语特征向量;
将所述视频特征向量、所述文本特征向量和所述词语特征向量进行融合,得到所述词语的中间向量。
3. 根据权利要求2所述的方法,其特征在于,所述将所述视频特征向量、所述文本特征向量和所述词语特征向量进行融合,得到所述词语的中间向量,包括:
将所述视频特征向量、所述文本特征向量和所述词语特征向量进行第一融合,得到第一融合向量;
将所述第一融合向量和所述词语特征向量进行第二融合,得到所述词语的中间向量。
4. 根据权利要求3所述的方法,其特征在于,所述将所述视频特征向量、所述文本特征向量和所述词语特征向量进行第一融合,得到第一融合向量,包括:
将所述视频特征向量、所述文本特征向量和所述词语特征向量进行依次拼接,得到第一拼接向量;
将所述第一拼接向量进行全连接特征映射,得到所述第一融合向量。
5. 根据权利要求3所述的方法,其特征在于,所述将所述第一融合向量和所述词语特征向量进行第二融合,得到所述词语的中间向量,包括:
将所述第一融合向量和所述词语特征向量进行依次拼接,得到第二拼接向量;
将所述第二拼接向量进行全连接特征映射,得到所述词语的中间向量。
6. 根据权利要求1所述的方法,其特征在于,所述基于所述词语的中间向量,生成所述词语的词权重,包括:
将所述中间向量进行维度变换,得到一维向量;
将所述一维向量通过阈值函数进行转换,得到所述词语的词权重。
7. 根据权利要求2至5任一所述的方法,其特征在于,所述提取所述视频的视频特征向量,包括:
对所述视频进行分帧操作,得到至少两个视频帧;
提取所述至少两个视频帧的视频帧向量;
计算所述至少两个视频帧的视频帧向量的平均向量,将所述平均向量确定为所述视频特征向量;或,计算所述至少两个视频帧的视频帧向量的加权向量,将所述加权向量确定为所述视频特征向量。
8. 根据权利要求7所述的方法,其特征在于,所述提取所述视频的视频特征向量,包括:
调用残差神经网络提取所述视频中的至少两个视频帧的视频帧向量。
9. 根据权利要求2至5任一所述的方法,其特征在于,所述提取所述视频关联文本的文

本特征向量,包括:

调用双向编码转换网络提取所述视频关联文本的文本特征向量;

或,

调用长短期记忆网络提取所述视频关联文本的文本特征向量。

10. 根据权利要求2至5任一所述的方法,其特征在于,所述提取所述词语的词语特征向量,包括:

对所述视频关联文本进行分词,得到所述词语;

调用深度神经网络提取所述词语的词语特征向量。

11. 根据权利要求10所述的方法,其特征在于,所述对所述视频关联文本进行分词,得到所述词语,包括:

调用中文分词工具对所述视频关联文本进行分词,得到所述词语。

12. 根据权利要求7所述的方法,其特征在于,所述提取所述视频的视频特征向量,包括:

基于所述视频中的所述视频帧,提取得到所述视频帧向量;

基于所述视频中的音频帧,提取得到音频帧向量;

基于所述视频帧中的文本,提取得到文本幕向量;

将所述视频帧向量、所述音频帧向量和所述文本幕向量中的至少两种进行融合,得到所述视频特征向量。

13. 一种词权重的生成装置,其特征在于,所述装置包括:

获取模块,用于获取视频和视频关联文本,所述视频关联文本包括至少一个词语;

生成模块,用于基于所述视频、所述视频关联文本和所述词语的多模态特征融合,生成所述词语的中间向量;

生成模块,还用于基于所述词语的中间向量,生成所述词语的词权重。

14. 一种计算机设备,其特征在于,所述计算机设备包括:处理器和存储器,所述存储器存储有计算机程序,所述计算机程序由所述处理器加载并执行以实现如权利要求1至12任一所述的词权重生成方法。

15. 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存储有计算机程序,所述计算机程序由处理器加载并执行以实现如权利要求1至12任一所述的词权重生成方法。

词权重的生成方法、装置、设备及介质

技术领域

[0001] 本申请涉及信息处理领域,特别涉及一种词权重的生成方法、装置、设备及介质。

背景技术

[0002] 视频标题是视频内容的主要成分之一。在视频搜索系统中需要基于视频语义信息的理解,对视频标题中的各个词语的权重值进行预先提取,以便于后续的视频搜索过程。

[0003] 相关技术中,对词权重的生成方法主要是对句子和词语进行分别编码,得到句子向量和词语向量。对编码完成后的句子向量和词语向量进行特征融合,得到融合向量,对融合向量进行二分类判断,判断当前词语是否为核心词,进而输出当前词语的词权重。

[0004] 上述方法生成的词权重在某些情况下是不准确的。比如,视频标题是“双击这个视频,你会发现红烧肉比酸菜鱼的做法更简单”,上述方法难以对“红烧肉”和“酸菜鱼”的权重做出有力区分。

发明内容

[0005] 本申请提供了一种词权重生成方法、装置、设备及介质,通过融入视频的画面特征信息,能够提高词权重值的准确率和可靠程度。所述技术方案如下:

[0006] 根据本申请的一个方面,提供了一种词权重生成方法,所述方法包括:

[0007] 获取视频和视频关联文本,视频关联文本包括至少一个词语;

[0008] 基于视频、视频关联文本和词语的多模态特征融合,生成词语的中间向量;

[0009] 基于词语的中间向量,生成词语的词权重。

[0010] 根据本申请的一个方面,提供了一种词权重的生成装置,所述装置包括:

[0011] 获取模块,用于获取视频和视频关联文本,视频关联文本包括至少一个词语;

[0012] 生成模块,用于基于视频、视频关联文本和词语的多模态特征融合,生成词语的中间向量;

[0013] 生成模块,还用于基于词语的中间向量,生成词语的词权重。

[0014] 在一个可选的实施例中,生成模块包括提取模块和融合模块。

[0015] 在一个可选的实施例中,提取模块,用于提取视频的视频特征向量;提取视频关联文本的文本特征向量;以及提取词语的词语特征向量;

[0016] 在一个可选的实施例中,融合模块,用于将视频特征向量、文本特征向量和词语特征向量进行融合,得到词语的中间向量。

[0017] 在一个可选的实施例中,融合模块包括第一融合子模块和第二融合子模块。

[0018] 在一个可选的实施例中,第一融合子模块,用于将视频特征向量、文本特征向量和词语特征向量进行第一融合,得到第一融合向量;

[0019] 在一个可选的实施例中,第二融合子模块,用于将第一融合向量和词语特征向量进行第二融合,得到词语的中间向量。

[0020] 在一个可选的实施例中,第一融合子模块包括第一拼接模块和第一映射模块。

- [0021] 在一个可选的实施例中,第一拼接模块,用于将视频特征向量、文本特征向量和词语特征向量进行依次拼接,得到第一拼接向量;
- [0022] 在一个可选的实施例中,第一映射模块,用于将第一拼接向量进行全连接特征映射,得到第一融合向量。
- [0023] 在一个可选的实施例中,第二融合子模块包括第二拼接模块和第二映射模块。
- [0024] 在一个可选的实施例中,第二拼接模块,用于将第一融合向量和词语特征向量进行依次拼接,得到第二拼接向量;
- [0025] 在一个可选的实施例中,第二映射模块,用于将第二拼接向量进行全连接特征映射,得到词语的中间向量。
- [0026] 在一个可选的实施例中,生成模块还包括转换模块。
- [0027] 在一个可选的实施例中,转换模块,用于将中间向量进行维度变换,得到一维向量;
- [0028] 在一个可选的实施例中,转换模块,还用于将一维向量通过阈值函数进行转换,得到词语的词权重。
- [0029] 在一个可选的实施例中,提取模块包括视频提取模块、文本提取模块和词语提取模块。其中,视频提取模块包括分帧模块、提取子模块和计算模块。
- [0030] 在一个可选的实施例中,分帧模块用于对视频进行分帧操作,得到至少两个视频帧;
- [0031] 在一个可选的实施例中,提取子模块用于提取至少两个视频帧的视频帧向量;
- [0032] 在一个可选的实施例中,计算模块用于计算至少两个视频帧的视频帧向量的平均向量,将平均向量确定为视频特征向量;或,计算至少两个视频帧的视频帧向量的加权向量,将加权向量确定为视频特征向量。
- [0033] 在一个可选的实施例中,提取子模块还用于调用残差神经网络提取视频中的至少两个视频帧的视频帧向量。
- [0034] 在一个可选的实施例中,文本提取模块,用于调用双向编码转换网络提取视频关联文本的文本特征向量,或,调用长短期记忆网络提取视频关联文本的文本特征向量。
- [0035] 在一个可选的实施例中,词语提取模块包括分词模块和词语提取子模块。
- [0036] 在一个可选的实施例中,分词模块,用于对视频关联文本进行分词,得到词语;
- [0037] 在一个可选的实施例中,词语提取子模块,用于调用深度神经网络提取词语的词语特征向量。
- [0038] 在一个可选的实施例中,分词模块还用于调用中文分词工具对视频关联文本进行分词,得到词语。
- [0039] 根据本申请的一个方面,提供了一种计算机设备,所述计算机设备包括:处理器和存储器,所述存储器存储有计算机程序,所述计算机程序由所述处理器加载并执行以实现如上所述的词权重生成方法。
- [0040] 根据本申请的另一方面,提供了一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序由处理器加载并执行以实现如上所述的词权重生成方法。
- [0041] 根据本申请的另一个方面,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

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词权重生成方法。

[0042] 本申请实施例提供的技术方案带来的有益效果至少包括:

[0043] 通过将视频、视频关联文本和词语进行多维度的特征结合,生成中间向量,基于中间向量,生成词语的词权重。在视频搜索过程中,采用上述词权重生成方法来预先提取词语的权重值,不仅考虑了文本维度的特征,还引入融合了视频维度的特征,基于多维度的特征来进行词权重生成,有利于提升输出的词权重的准确率和可靠程度,提高了视频关联文本中对关键词语和混淆词语之间的区分度。

附图说明

[0044] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例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申请的一些实施例,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他的附图。

[0045] 图1是根据一示例性实施例提供的一种词权重生成系统的示意图;

[0046] 图2是本申请一个示例性实施例提供的词权重的生成方法的流程图;

[0047] 图3是本申请一个示例性实施例提供的词权重生成模型的示意图;

[0048] 图4是本申请另一个示例性实施例提供的词权重生成模型的示意图;

[0049] 图5是本申请一个示例性实施例提供的词权重生成方法的流程图;

[0050] 图6是本申请一个示例性实施例提供的生成视频特征向量的流程图;

[0051] 图7是本申请一个示例性实施例提供的生成文本特征向量的流程图;

[0052] 图8是本申请一个示例性实施例提供的生成词语特征向量的流程图;

[0053] 图9是本申请一个示例性实施例提供的训练词权重生成模型的流程图;

[0054] 图10是本申请一个示例性实施例提供的词权重生成方法流程图;

[0055] 图11是本申请的一个示例性实施例提供的词权重生成装置的结构框图;

[0056] 图12示出了本申请一个示例性实施例提供的电子设备的结构框图;

[0057] 图13示出了本申请一个示例性实施例提供的数据共享系统的示意图;

[0058] 图14示出了本申请一个示例性实施例提供的区块链结构的示意图;

[0059] 图15示出了本申请一个示例性实施例提供的新区块生成过程的示意图。

具体实施方式

[0060] 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。

[0061] 首先,对本申请实施例中涉及的名词进行简单介绍:

[0062] **词语重要度**:指词语对句子表达的意思起到的作用大小。常见的,一个完整句子的组成部分包括主语、谓语、宾语、定语、状语和补语,示例性的,句子为“双击这个视频,你会发现红烧肉比酸菜鱼的做法更简单”,去掉连接词和人称代词,该句子主要由词语“双击”“视频”“发现”“红烧肉”“酸菜鱼”“做法”“简单”组成。基于对句子整体的意思理解,容易得到“红烧肉”、“酸菜鱼”对句子的意思表达起到关键作用。更为具体的,“红烧肉”比“酸菜鱼”

对句子意思表达起的作用更进一步,即,“红烧肉”词语的重要度比“酸菜鱼”词语的重要度高。

[0063] 在一个实施例中,采用词语在句子中的权重值来表示词语的重要度。示意性的,上述句子中,“红烧肉”的权重值为0.91,“酸菜鱼”的权重值为0.82,即通过权重值的大小比较,可得“红烧肉”比“酸菜鱼”更重要。

[0064] 残差神经网络(Residual Network, ResNet):一种基于深度学习的特征提取神经网络。在传统的深度学习中,在神经网络可以收敛的前提下,随着网络深度增加,网络的表现先是逐渐增加至饱和,然后迅速下降,即为网络退化问题;在传统的深度学习中,存在梯度弥散问题;残差神经网络为深度学习神经网络添加一个恒等映射,解决了上述网络退化问题和梯度弥散问题。

[0065] 在本申请中,残差神经网络用于将自然语言转换为可进行运算的数学语言,示意性的,残差神经网络将视频帧转换为视频帧向量,该视频帧向量包含了视频帧的信息,即可用该视频帧向量替换上述视频帧。

[0066] 双向编码转换模型(Bidirectional Encoder Representations from Transformers, BERT):一种句子转换模型,可实现将真实世界抽象存在的文字转换成能够进行数学公式操作的向量。在一个实施例中,BERT将输入的文本转换为文本向量,该文本向量包含文本的信息,即可用该文本向量替换上述文本。

[0067] 深度神经网络(Deep Neural Networks, DNN):含有全连接的神经元结构的多层神经网络,实现把真实世界存在的客观事物转换为可以进行数学公式操作的向量。在一个实施例中,DNN将输入的词语转换为词向量,该词向量包含词语的信息,即可用该词向量替换上述词语。

[0068] 阈值函数:实现数值区间的转换,例如,数字x所处区间为 $[0, 100]$,通过阈值函数,将数字x转换为区间 $[0, 1]$ 的数字y。sigmoid(一种阈值函数)实现将一维向量映射为区间 $[0, 1]$ 上的数字,在本申请中,通过将一维向量映射到区间 $[0, 1]$ 上,得到词权重。

[0069] 云计算(cloud computing)指IT(Internet Technology, 互联网技术)基础设施的交付和使用模式,指通过网络以按需、易扩展的方式获得所需资源;广义云计算指服务的交付和使用模式,指通过网络以按需、易扩展的方式获得所需服务。这种服务可以是IT和软件、互联网相关,也可是其他服务。云计算是网格计算(Grid Computing)、分布式计算(Distributed Computing)、并行计算(Parallel Computing)、效用计算(Utility Computing)、网络存储(Network Storage Technologies)、虚拟化(Virtualization)、负载均衡(Load Balance)等传统计算机和网络技术发展融合的产物。

[0070] 随着互联网、实时数据流、连接设备多样化的发展,以及搜索服务、社会网络、移动商务和开放协作等需求的推动,云计算迅速发展起来。不同于以往的并行分布式计算,云计算的产生从理念上将推动整个互联网模式、企业管理模式发生革命性的变革。

[0071] 本申请实施例的方案包括模型训练阶段和词权重预测阶段。图1是根据一示例性实施例示出的一种词权重生成系统的示意图。如图1所示,在模型训练阶段,模型训练设备110通过预先设置好的训练样本集训练出准确性较高的词权重生成模型,在词权重预测阶段,词权重生成设备120根据训练出的词权重生成模型以及输入的视频和文本,预测文本中词语的权重值。

[0072] 其中,上述模型训练设备110和词权重预测设备120可以是具有机器学习能力的计算机设备,比如,该计算机设备可以是终端或服务器。

[0073] 可选的,上述模型训练设备110和词权重预测设备120可以是同一个计算机设备,或者,模型训练设备110和词权重预测设备120也可以是不同的计算机设备。并且,当模型训练设备110和词权重预测设备120是不同的设备时,模型训练设备110和词权重预测设备120可以是同一类型的设备,比如模型训练设备110和词权重预测设备120可以都是服务器;或者,模型训练设备110和词权重预测设备120也可以是不同类型的设备。上述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、内容分发网络(Content Delivery Network,CDN)、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。上述终端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等,但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接,本申请在此不做限制。

[0074] 针对词权重的预测阶段:

[0075] 为提升生成的词权重值的准确率和可靠程度,采用图2所述的方法生成词语的词权重。图2示出了本申请一个示例性实施例提供的词权重的生成方法流程图,如图2所示,该方法包括:

[0076] 步骤201:获取视频和视频关联文本,视频关联文本包括至少一个词语;

[0077] 视频和视频关联文本之间存在对应关系。

[0078] 在一个实施例中,视频关联文本是与视频相对应的标题,视频关联文本与视频彼此独立,其中,该标题由人为标注或机器生成,用于简要阐述该视频的中心含义。

[0079] 在一个实施例中,视频关联文本是与视频相对应的视频简介,视频关联文本与视频彼此独立,其中,该视频简介由人为撰写或机器生成,用于简要阐述该视频的具体内容。

[0080] 步骤202:基于视频、视频关联文本和词语的多模态特征融合,生成词语的中间向量;

[0081] 其中,多模态特征融合指对视频、视频关联文本、词语分别进行特征提取,得到视频特征向量、文本特征向量和词语特征向量,之后对视频特征向量、文本特征向量和词语特征向量进行向量融合运算。

[0082] 上述多模态特征融合包括以下两个步骤:

[0083] 第一,提取视频的视频特征向量;提取视频关联文本的文本特征向量;以及提取词语的词语特征向量;

[0084] 其中,视频特征向量指示视频特征信息、文本特征向量指示视频关联文本的特征信息、词语特征向量指示词语特征信息。

[0085] 第二,将视频特征向量、文本特征向量和词语特征向量进行融合,得到词语的中间向量。

[0086] 其中,中间向量包含视频、视频关联文本和词语的特征信息。

[0087] 步骤203:基于词语的中间向量,生成词语的词权重。

[0088] 示例性的,基于词语的中间向量生成词语的词权重包括以下两个步骤:

[0089] 第一、将中间向量进行维度变换,得到一维向量;

[0090] 响应于对视频、视频关联文本和词语进行多模态特征融合,生成词语的中间向量,其中,中间向量为包含视频、视频关联文本和词语的特征信息的多维度向量;在一个实施例中,通过全连接映射实现中间向量与一维向量的变换。

[0091] 第二、将一维向量通过阈值函数进行转换,得到词语的词权重。

[0092] 在一个实施例中,通过上述全连接映射对中间向量进行维度变换,如中间向量维度为388维,进行维度变换得到1维向量。其中,1维向量包含了词向量在句子向量的重要度信息。在一个实施例中,sigmoid对1维向量进行数值区间的转换,通过将一维向量映射到区间[0,1]上,得到词权重。

[0093] 综上所述,通过将视频、视频关联文本和词语的特征结合,生成中间向量,基于中间向量,生成词语的词权重。在视频搜索过程中,采用上述词权重生成方法来预先提取词语的权重值,不仅考虑了文本维度的特征,还引入融合了视频维度的特征,基于多维度的特征来进行词权重生成,有利于提升输出的词权重的准确率和可靠程度,提高了视频关联文本中对关键词语和混淆词语之间的区分度。

[0094] 图3示出了本申请一个示例性实施例提供的词权重生成模型的结构框图,图3中词权重生成模型300包括:分词网络310、转换网络320、融合网络330和映射网络340。其中,分词网络310用于将视频关联文本进行分词得到至少一个词语;转换网络320用于将视频转换为视频特征向量、将视频关联文本转化为文本特征向量、将词语转换为词语特征向量;融合网络330用于将视频特征向量、文本特征向量和词语特征向量融合得到中间向量;射网络340用于将中间向量映射得到词权重。

[0095] 图4示出了本申请另一个示例性实施例的词权重生成模型。词权重生成模型包括分词网络310、转换网络320、融合网络330、映射网络340。转换网络320包括第一转换子网络321、第二转换子网322和第三转换子网络323。融合网络330包括第一融合子网络331和第二融合子网络332。

[0096] 图5示出了本申请一个示例性实施例的词权重生成方法流程图。结合参考图4的词权重生成模型,该词权重生成方法包括:

[0097] 步骤510:调用第一转换子网络321对视频进行处理,输出视频的视频特征向量;

[0098] 示例性的,第一转换子网络321用于对视频进行分帧操作,得到至少两个视频帧,然后提取至少两个视频帧的视频帧向量,再计算至少两个视频帧的视频帧向量的平均向量,将平均向量确定为视频特征向量;或,计算至少两个视频帧的视频帧向量的加权向量,将加权向量确定为视频特征向量。

[0099] 可选的,上述分帧操作至少包括以下两种处理方式:

[0100] 第一、根据固定时间间隔提取视频帧;

[0101] 示意性的,假设视频时长为30s,预设采样时长间隔为0.2s,则对视频进行分帧操作指每隔0.2s采集视频帧。

[0102] 第二、根据预设的采集规则提取视频帧。

[0103] 在一个实施例中,假设视频时长为30s,预先设定在视频时长的前20%时长内,每隔1s采集视频帧,在视频时长的中间60%时长内,每隔0.2s采集视频帧,在视频时长的后20%时长内,每隔1s采集视频帧。

[0104] 可选的,上述提取至少两个视频帧的视频帧向量包括:调用残差神经网络提取视

频中的至少两个视频帧的视频帧向量。

[0105] 示意性的,如图6所示,视频601分帧得到四个视频帧602,四个视频帧602输入至ResNet603,分别得到第一帧向量、第二帧向量、第三帧向量和第四帧向量。将上述四个帧向量取平均或加权得到视频帧向量。在一个实施例中,上述计算至少两个视频帧的视频帧向量的平均向量,指对第一帧向量、第二帧向量、第三帧向量和第四帧向量进行累加之后求平均值。在一个实施例中,上述计算至少两个视频帧的视频帧向量的加权向量,指对第一帧向量、第二帧向量、第三帧向量和第四帧向量进行加权求和。例如,第一帧向量为a、第二帧向量为b、第三帧向量为c和第四帧向量为d,假设对第一帧向量赋予权重0.3,第二帧向量赋予权重0.1,第三帧向量赋予权重0.2和第四帧向量赋予权重0.4,则得到的视频特征向量为 $0.3a+0.1b+0.2c+0.4d$ 。

[0106] 步骤520:调用第二转换子网络322对视频关联文本进行处理,输出视频关联文本的文本特征向量;

[0107] 在一个实施例中,第二转换子网络322包括双向编码转换网络或长短期记忆网络。调用双向编码转换网络提取视频关联文本的文本特征向量,或,调用长短期记忆网络提取所述视频关联文本的文本特征向量。

[0108] 可选的,如图7所示,视频关联文本701输入至Bert网络702,得到文本特征向量。

[0109] 步骤530:调用分词网络310对视频关联文本进行分词,得到词语;

[0110] 在一个可选的实施例中,分词网络内设jieba(一种第三方中文分词库),jieba内支持三种分词模式,第一、精确模式:将语句进行最精确的切分,不存在冗余数据,适合做文本分析;第二、全模式:将语句中所有可能是词的词语都切分出来,速度很快,但是存在冗余数据;第三、搜索引擎模式:在精确模式的基础上,对长词再次进行切分。在实际使用场景中,根据视频关联文本的类型、长短等对模式进行选择,最终实现将视频关联文本转换为至少一个词语。

[0111] 在一个实施例中,视频关联文本为“这鲁班没救了,经济被压制,完全起不来,手机给你来玩!”,其中,在精确模式下分词得到的词语包括“这”“鲁班”“没救了”“经济”“被”“压制”“完全”“起不来”“手机”“给”“你来玩”“!”;在全模式下分词得到的词语包括“鲁班”“没救”“经济”“压制”“完全”“手机”;在搜索引擎模式下分词得到的词语包括“这”“鲁班”“没”“救”“了”“经济”“被”“压制”“完全”“起”“不”“来”“手机”“给”“你”“来”“玩”“!”

[0112] 步骤540:调用第三转换子网络323对词语进行处理,输出词语的词语特征向量;

[0113] 在一个实施例中,第三转换子网络323包括深度神经网络。调用深度神经网络提取词语的词语特征向量。示意性的,如图8所示,词语输入DNN801,得到词语向量。

[0114] 步骤550:调用第一融合子网络331将视频特征向量、文本特征向量和词语特征向量进行第一融合,得到第一融合向量;

[0115] 在一个实施例中,调用第一融合子网络将视频特征向量、文本特征向量和词语特征向量进行依次拼接,得到第一拼接向量,将第一拼接向量进行全连接特征映射,得到第一融合向量。

[0116] 上述拼接指对所有向量进行维度拼接,如原本视频帧向量维度为318维、文本向量为50维、词向量为10维,则得到的第一拼接向量维度为378维。在一个实施例中,上述全连接特征映射指对得到的第一拼接向量进行映射,得到第一融合向量。示意性的,第一拼接向量

为 $[a, b, c]$,其中 a, b, c 分别指示视频信息、视频关联文本信息和词语信息,通过全连接层映射得到第一融合向量 $[0.9a, 3b, 10c]$,其中, $0.9a, 3b, 10c$ 、分别指示视频信息、视频关联文本信息和词语信息,即全连接特征映射改变了视频信息、视频关联文本信息和词语信息之间的融合程度。上述示例仅起到解释说明作用,实际全连接特征映射实现在高维空间,且融合的程度随着输入的视频、视频关联文本和词语的改变随之发生变化。

[0117] 步骤560:调用第二融合子网络332将第一融合向量和词语特征向量进行第二融合,得到词语的中间向量;

[0118] 在一个实施例中,调用第二融合子网络332将第一融合向量和词语特征向量进行依次拼接,得到第二拼接向量,将第二拼接向量进行全连接特征映射,得到词语的中间向量。拼接和全连接特征映射同第一融合子网络相类似,不再赘述。通过上述第一融合子网络的拼接和第二融合子网络的拼接,强化了当前词语的重要性,提升了词语特征向量在中间向量的权重。

[0119] 步骤570:调用映射网络340将中间向量进行维度变换,得到一维向量,并将一维向量通过阈值函数进行转换,得到词语的词权重。

[0120] 在一个实施例中,通过上述全连接映射对中间向量进行维度变换,如中间向量维度为388维,进行维度变换得到1维向量。其中,1维向量包含了词语特征向量在文本特征向量的重要度信息。在一个实施例中,采用sigmoid对1维向量进行数值区间的转换,通过将一维向量映射到区间 $[0, 1]$ 上得到词权重。

[0121] 综上所述,本实施例提供的方法,通过对视频、视频关联文本和词语进行特征提取得到视频特征向量、文本特征向量和词语特征向量,再将上述三种模态的特征向量进行拼接和全连接映射,得到第一融合向量,然后将第一融合向量和词语特征向量进行拼接和全连接映射,得到中间向量,基于中间向量,得到词语的词权重。

[0122] 本实施例提供的方法,还通过对词语特征向量进行了两次拼接,强化了当前词语特征向量在中间向量的信息量占比,有利于提高视频关联文本中不同词语的权重值区分度。

[0123] 本实施例提供的方法,还对视频特征向量、文本特征向量和词语特征向量进行融合,不仅考虑了文本维度的特征,还引入融合了视频维度的特征,基于多维度的特征来进行词权重生成,有利于提升输出的词权重的准确率和可靠程度,提高了视频关联文本中对关键词语和混淆词语之间的区分度。

[0124] 本实施例提供的方法,还通过采用残差神经网络对视频进行特征提取、采用双向编码转换网络或长短期记忆网络对视频关联文本进行特征提取和采用深度神经网络对词语进行特征提取,实现了将自然语言转换为可进行数学运算的特征向量,简化了本申请词权重生成方法的数学运算。

[0125] 上述词权重生成模型是采用训练方法训练得到的。图9是本申请的一个示例性实施例的图像修复模型的训练方法流程图。

[0126] 步骤901:获取样本视频、样本视频关联文本和样本词权重;

[0127] 样本视频和样本视频关联文本之间存在对应关系。

[0128] 步骤902:将样本视频、样本视频关联文本输入词权重生成模型;

[0129] 步骤903:获取词权重生成模型输出的预测词权重;

[0130] 预测词权重指的是样本视频和样本视频关联文本输入词权重生成模型输出的词权重。

[0131] 步骤904:计算样本词权重和预测词权重的误差;

[0132] 步骤905:根据误差,优化词权重生成模型的网络参数。

[0133] 词权重生成模型的网络参数用于调整词权重生成模型的性能,在本申请中,词权重生成模型的网络参数至少包括ResNet的网络参数、BERT的网络参数、DNN的网络参数,视频特征向量、文本特征向量和词语特征向量之间的融合参数。

[0134] 基于图2的可选实施例中,步骤201获取视频和视频关联文本,视频关联文本包括至少一个词语中,视频的获取方法包括:逐个获取目标视频库中的视频文件作为目标视频进行后续处理。

[0135] 在一个实施例中,目标视频为视频库中已存储视频文件的一个视频片段,该目标视频的提取包括如下方式中的至少一种:

[0136] (1) 基于预设时长区间对视频文件进行划分,如:提取视频文件开头前两分钟的视频片段作为视频。

[0137] (2) 通过人工手动对目标视频进行提取。

[0138] 也即,基于人工对视频库中已存储的视频文件进行提取,如:观看人认为视频文件中第5至第6分钟的视频片段为本视频文件的核心视频,观看人提取该核心视频作为目标视频。

[0139] (3) 通过视频提取模型对目标视频进行提取。

[0140] 即,将视频库中存储的视频文件输入视频提取模型,由视频提取模型对上述视频文件进行特征提取后,对上述视频文件中的帧与帧之间的关联性进行分析,从而对视频文件进行提取,得到目标视频。

[0141] 在一个实施例中,目标视频为符合筛选条件的视频文件,示意性的,目标视频为指定用户上传的视频文件,或,目标视频为符合要求的视频类型的视频文件,或,视频时长达到阈值的视频文件。

[0142] 针对目标视频为符合要求的视频类型的视频文件的情况,示例性的,当视频文件为电视剧中的某一集、电影视频、电影片段、纪录片视频等类型的视频时,将该视频文件作为目标视频进行获取。

[0143] 针对上述目标视频为指定用户上传的视频文件,示例性的,当视频文件为某专业机构上传的视频、某公共人物上传的视频、某权威人士上传的视频时,将该视频文件作为目标视频进行获取。

[0144] 基于图2所示的可选实施例中,提取视频的视频特征向量包括以下步骤:

[0145] 第一、基于视频中的视频帧,提取得到视频帧向量;

[0146] 其中,视频包括视频帧和音频帧,此处的视频帧表现为画面。其中,画面特征是指从视频的界面表现上提取得到的特征,其中,画面特征中包括与主题名称、弹幕、对白等文本内容对应的特征,也包括与视频画面帧对应的特征。

[0147] 在一个可选的实施例中,采用ResNet特征提取视频帧得到视频帧向量,即将视频帧由原本的自然语言转换为能进行数学运算的向量。

[0148] 第二、基于视频中的音频帧,提取得到音频帧向量;

[0149] 音频帧表现为视频中的声音,在一个实施例中,音频帧与画面帧之间达成匹配,即音画同步,即在同一时间点同时提取音频帧和画面帧;在一个实施例中,音频帧与画面帧之间不匹配,即音画异步,即提取音频帧和画面帧的时间点不一致。

[0150] 在一个可选的实施例中,采用卷积神经网络(Convolutional Neural Networks, CNN)特征提取音频帧,得到音频帧向量,即将音频帧由原本的自然语言转换为能进行数学运算的向量。

[0151] 第三、基于视频帧中的文本,提取得到文本幕向量;

[0152] 视频帧中的文本是指与目标视频相关的,目标视频所涉及的文本内容,示意性的,视频帧中的文本包括弹幕内容、画面中出现的内容、对白内容等。

[0153] 在一个可选的实施例中,采用BERT特征提取画面上的文本,得到文本幕向量,即将画面上的文本由原本的自然语言转换为能进行数学运算的向量。

[0154] 第四、将所述视频帧向量、所述音频帧向量和所述文本幕向量中的至少两种进行融合,得到所述视频特征向量。

[0155] 在一个实施例中,采用加权方式实现视频帧向量、音频帧向量和文本幕向量中的至少两种进行融合,示意性的,视频帧向量为 x 、音频帧向量为 y 、文本幕向量为 z ,假设对视频帧向量赋予权重0.5,音频帧向量赋予权重0.1,文本幕向量赋予权重0.4,则得到的视频特征向量为 $0.5x+0.1y+0.4z$ 。

[0156] 图10是本申请一个示例性实施例提供的词权重生成方法的流程图。示例性的,输入句子“双击这个视频,你会发现红烧肉比酸菜鱼的做法更简单”可表示为 $text=[x_0, x_1, \dots]$,输入词语为 x_i ,抽取得到的视频关键帧为 f_i ,则句子的编码向量为 $V_{text}=BERT(text)$,关键帧编码向量为 $V_{img}=ResNet(f_i)$,词语的编码向量为 $V_{word}=DNN(x_i)$,则第一融合向量 $Fusion1=fusion(V_{text}, V_{img}, V_{word})$,其中 $fusion$ 为多类特征向量拼接后通过全连接方式完成特征映射得到。第二次融合的输入为首次融合得到的 $Fusion1$ 向量和词语向量,第二融合向量 $Fusion2=fusion(Fusion1, V_{word})$,模型特征融合过程中的两次融合强化了词语的重要性,可以有效的识别该词在句子的重要程度,即词权重值。图10中用“○”指示关键帧编码向量一个维度信息,用“●”指示句子编码向量的一个维度信息,用“⊙”指示词语编码向量的一个维度信息。第一次融合向量 $Fusion1$ 和第二次融合向量 $Fusion2$ 采用上述三种圆的占比关系来表示关键帧编码向量、句子编码向量和词语编码向量的融合程度。

[0157] 应用场景:

[0158] 在一个实施例中,当用户在终端上进行视频搜索时,服务端对视频库内的视频进行预处理,根据视频和视频标题得到该视频的核心词,并根据得到的核心词判断是否在终端上展示该视频。

[0159] 上述预处理包括以下过程:服务器首先获取视频和视频标题,视频标题包括至少一个词语;其中,视频标题与视频彼此独立,其中,该标题由人为标注或机器生成,用于简要阐述该视频的中心含义。

[0160] 然后,基于视频、视频标题和词语的多模态特征融合,服务器生成词语的中间向量;其中,多模态特征融合指对视频、视频标题、词语分别进行特征提取,得到视频特征向量、标题文本特征向量和词语特征向量,之后对视频特征向量、文本特征向量和词语特征向

量进行向量融合运算。具体特征融合过程参考上述图5所示的实施例所示的细节。

[0161] 最后,基于词语的中间向量,服务器生成词语的词权重。示意性的,服务器将中间向量进行维度变换,得到一维向量,再将一维向量通过阈值函数进行转换,得到词语的词权重。

[0162] 图11是本申请一个示例性实施例的词权重生成装置的结构框图,该装置包括:

[0163] 获取模块1120,用于获取视频和视频关联文本,视频关联文本包括至少一个词语;

[0164] 生成模块1140,用于基于视频、视频关联文本和词语的多模态特征融合,生成词语的中间向量;

[0165] 生成模块1140,还用于基于词语的中间向量,生成词语的词权重。

[0166] 在一个可选的实施例中,生成模块1140包括提取模块41和融合模块42:

[0167] 在一个可选的实施例中,提取模块41,用于提取视频的视频特征向量;提取视频关联文本的文本特征向量;以及提取词语的词语特征向量;

[0168] 在一个可选的实施例中,融合模块42,用于将视频特征向量、文本特征向量和词语特征向量进行融合,得到词语的中间向量。

[0169] 在一个可选的实施例中,融合模块42包括第一融合子模块421和第二融合子模块422。

[0170] 在一个可选的实施例中,第一融合子模块421,用于将视频特征向量、文本特征向量和词语特征向量进行第一融合,得到第一融合向量;

[0171] 在一个可选的实施例中,第二融合子模块422,用于将第一融合向量和词语特征向量进行第二融合,得到词语的中间向量。

[0172] 在一个可选的实施例中,第一融合子模块421包括第一拼接模块211和第一映射模块212。

[0173] 在一个可选的实施例中,第一拼接模块211,用于将视频特征向量、文本特征向量和词语特征向量进行依次拼接,得到第一拼接向量;

[0174] 在一个可选的实施例中,第一映射模块212,用于将第一拼接向量进行全连接特征映射,得到第一融合向量。

[0175] 在一个可选的实施例中,第二融合子模块422包括第二拼接模块221和第二映射模块222。

[0176] 在一个可选的实施例中,第二拼接模块221,用于将第一融合向量和词语特征向量进行依次拼接,得到第二拼接向量;

[0177] 在一个可选的实施例中,第二映射模块222,用于将第二拼接向量进行全连接特征映射,得到词语的中间向量。

[0178] 在一个可选的实施例中,生成模块1140还包括转换模块43。

[0179] 在一个可选的实施例中,转换模块43,用于将中间向量进行维度变换,得到一维向量;

[0180] 在一个可选的实施例中,转换模块43,还用于将一维向量通过阈值函数进行转换,得到词语的词权重。

[0181] 在一个可选的实施例中,提取模块包括视频提取模块411、文本提取模块412和词语提取模块413。其中,视频提取模块411包括分帧模块111、提取子模块112和计算模块113。

[0182] 在一个可选的实施例中,分帧模块111用于对视频进行分帧操作,得到至少两个视频帧;

[0183] 在一个可选的实施例中,提取子模块112用于提取至少两个视频帧的视频帧向量;

[0184] 在一个可选的实施例中,计算模块113用于计算至少两个视频帧的视频帧向量的平均向量,将平均向量确定为视频特征向量;或,计算至少两个视频帧的视频帧向量的加权向量,将加权向量确定为视频特征向量。

[0185] 在一个可选的实施例中,提取子模块112还用于调用残差神经网络提取视频中的至少两个视频帧的视频帧向量。

[0186] 在一个可选的实施例中,文本提取模块412,用于调用双向编码转换网络提取视频关联文本的文本特征向量,或,调用长短期记忆网络提取视频关联文本的文本特征向量。

[0187] 在一个可选的实施例中,词语提取模块413包括分词模块131和词语提取子模块132。

[0188] 在一个可选的实施例中,分词模块131,用于对视频关联文本进行分词,得到词语;

[0189] 在一个可选的实施例中,词语提取子模块132,用于调用深度神经网络提取词语的词语特征向量。

[0190] 在一个可选的实施例中,分词模块131还用于调用中文分词工具对视频关联文本进行分词,得到词语。

[0191] 综上所述,本装置通过对视频、视频关联文本和词语进行特征提取得到视频特征向量、文本特征向量和词语特征向量,再将上述三种模态的特征向量进行拼接和全连接映射,得到第一融合向量,然后将第一融合向量和词语特征向量进行拼接和全连接映射,得到中间向量,基于中间向量,得到词语的词权重。

[0192] 上述装置对词语特征向量进行了两次拼接,强化了当前词语特征向量在中间向量的信息量占比,有利于提高视频关联文本中不同词语的权重值区分度。

[0193] 上述装置实现了在视频搜索过程中,采用上述词权重生成方法来预先提取词语的权重值,不仅考虑了文本维度的特征,还引入融合了视频维度的特征,基于多维度的特征来进行词权重生成,有利于提升输出的词权重的准确率和可靠程度,提高了视频关联文本中对关键词语和混淆词语之间的区分度。

[0194] 上述装置采用残差神经网络对视频进行特征提取、采用双向编码转换网络或长短期记忆网络对视频关联文本进行特征提取和采用深度神经网络对词语进行特征提取,实现了将自然语言转换为可进行数学运算的特征向量,简化了本申请词权重生成装置的数学运算。

[0195] 图12示出了本申请一个示例性实施例提供的电子设备1200的结构框图。该电子设备1200可以是便携式移动终端,比如:智能手机、平板电脑、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、笔记本电脑或台式电脑。电子设备1200还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

[0196] 通常,电子设备1200包括有:处理器1201和存储器1202。

[0197] 处理器1201可以包括一个或多个处理核心,比如4核心处理器、8核心处理器等。处

处理器1201可以采用DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field-Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。处理器1201也可以包括主处理器和协处理器, 主处理器是用于对在唤醒状态下的数据进行处理的处理单元, 也称CPU (Central Processing Unit, 中央处理器); 协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中, 处理器1201可以集成有GPU (Graphics Processing Unit, 图像处理器), GPU用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中, 处理器1201还可以包括AI (Artificial Intelligence, 人工智能) 处理器, 该AI处理器用于处理有关机器学习的计算操作。

[0198] 存储器1202可以包括一个或多个计算机可读存储介质, 该计算机可读存储介质可以是非暂态的。存储器1202还可包括高速随机存取存储器, 以及非易失性存储器, 比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中, 存储器1202中的非暂态的计算机可读存储介质用于存储至少一个指令, 该至少一个指令用于被处理器1201所执行以实现本申请中方法实施例提供的图像修复方法。

[0199] 在一些实施例中, 电子设备1200还可选包括有: 外围设备接口1203和至少一个外围设备。处理器1201、存储器1202和外围设备接口1203之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口1203相连。具体地, 外围设备包括: 射频电路1204、显示屏1205、摄像头组件1206、音频电路1207、定位组件1208和电源1209中的至少一种。

[0200] 外围设备接口1203可被用于将I/O (Input/Output, 输入/输出) 相关的至少一个外围设备连接到处理器1201和存储器1202。在一些实施例中, 处理器1201、存储器1202和外围设备接口1203被集成在同一芯片或电路板上; 在一些其他实施例中, 处理器1201、存储器1202和外围设备接口1203中的任意一个或两个可以在单独的芯片或电路板上实现, 本实施例对此不加以限定。

[0201] 射频电路1204用于接收和发射RF (Radio Frequency, 射频) 信号, 也称电磁信号。射频电路1204通过电磁信号与通信网络以及其他通信设备进行通信。射频电路1204将电信号转换为电磁信号进行发送, 或者, 将接收到的电磁信号转换为电信号。可选地, 射频电路1204包括: 天线系统、RF收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路1204可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于: 万维网、城域网、内联网、各代移动通信网络 (2G、3G、4G及5G)、无线局域网和/或WiFi (Wireless Fidelity, 无线保真) 网络。在一些实施例中, 射频电路1204还可以包括NFC (Near Field Communication, 近距离无线通信) 有关的电路, 本申请对此不加以限定。

[0202] 显示屏1205用于显示UI (User Interface, 用户界面)。该UI可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏1205是触摸显示屏时, 显示屏1205还具有采集在显示屏1205的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器1201进行处理。此时, 显示屏1205还可以用于提供虚拟按钮和/或虚拟键盘, 也称软按钮和/或软键盘。在一些实施例中, 显示屏1205可以为一个, 设置在电子设备1200的前面板; 在另一些实施例中, 显示屏1205可以为至少两个, 分别设置在电子设备1200的不同表面或

呈折叠设计;在另一些实施例中,显示屏1205可以是柔性显示屏,设置在电子设备1200的弯曲表面上或折叠面上。甚至,显示屏1205还可以设置成非矩形的不规则图形,也即异形屏。显示屏1205可以采用LCD(Liquid Crystal Display,液晶显示屏)、OLED(Organic Light-Emitting Diode,有机发光二极管)等材质制备。

[0203] 摄像头组件1206用于采集图像或视频。可选地,摄像头组件1206包括前置摄像头和后置摄像头。通常,前置摄像头设置在终端的前面板,后置摄像头设置在终端的背面。在一些实施例中,后置摄像头为至少两个,分别为主摄像头、景深摄像头、广角摄像头、长焦摄像头中的任意一种,以实现主摄像头和景深摄像头融合实现背景虚化功能、主摄像头和广角摄像头融合实现全景拍摄以及VR(Virtual Reality,虚拟现实)拍摄功能或者其它融合拍摄功能。在一些实施例中,摄像头组件1206还可以包括闪光灯。闪光灯可以是单色温闪光灯,也可以是双色温闪光灯。双色温闪光灯是指暖光闪光灯和冷光闪光灯的组合,可以用于不同色温下的光线补偿。

[0204] 音频电路1207可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波,并将声波转换为电信号输入至处理器1201进行处理,或者输入至射频电路1204以实现语音通信。出于立体声采集或降噪的目的,麦克风可以为多个,分别设置在电子设备1200的不同部位。麦克风还可以是阵列麦克风或全向采集型麦克风。扬声器则用于将来自处理器1201或射频电路1204的电信号转换为声波。扬声器可以是传统的薄膜扬声器,也可以是压电陶瓷扬声器。当扬声器是压电陶瓷扬声器时,不仅可以将电信号转换为人类可听见的声波,也可以将电信号转换为人类听不见的声波以进行测距等用途。在一些实施例中,音频电路1207还可以包括耳机插孔。

[0205] 定位组件1208用于定位电子设备1200的当前地理位置,以实现导航或LBS(Location Based Service,基于位置的服务)。定位组件1208可以是基于美国的GPS(Global Positioning System,全球定位系统)、中国的北斗系统或俄罗斯的伽利略系统的定位组件。

[0206] 电源1209用于为电子设备1200中的各个组件进行供电。电源1209可以是交流电、直流电、一次性电池或可充电电池。当电源1209包括可充电电池时,该可充电电池可以是有线充电电池或无线充电电池。有线充电电池是通过有线线路充电的电池,无线充电电池是通过无线线圈充电的电池。该可充电电池还可以用于支持快充技术。

[0207] 在一些实施例中,电子设备1200还包括有一个或多个传感器1210。该一个或多个传感器1210包括但不限于:加速度传感器1211、陀螺仪传感器1212、压力传感器1213、指纹传感器1214、光学传感器1215以及接近传感器1216。

[0208] 加速度传感器1212可以检测以电子设备1200建立的坐标系的三个坐标轴上的加速度大小。比如,加速度传感器1211可以用于检测重力加速度在三个坐标轴上的分量。处理器1201可以根据加速度传感器1211采集的重力加速度信号,控制显示屏1205以横向视图或纵向视图进行用户界面的显示。加速度传感器1212还可以用于游戏或者用户的运动数据的采集。

[0209] 陀螺仪传感器1212可以检测电子设备1200的机体方向及转动角度,陀螺仪传感器1212可以与加速度传感器1211协同采集用户对电子设备1200的3D动作。处理器1201根据陀螺仪传感器1212采集的数据,可以实现如下功能:动作感应(比如根据用户的倾斜操作来改

变UI)、拍摄时的图像稳定、游戏控制以及惯性导航。

[0210] 压力传感器1213以设置在电子设备1200的侧边框和/或显示屏1205的下层。当压力传感器1213设置在电子设备1200的侧边框时,可以检测用户对电子设备1200的握持信号,由处理器1201根据压力传感器1213采集的握持信号进行左右手识别或快捷操作。当压力传感器1213设置在显示屏1205的下层时,由处理器1201根据用户对显示屏1205的压力操作,实现对UI界面上的可操作性控件进行控制。可操作性控件包括按钮控件、滚动条控件、图标控件、菜单控件中的至少一种。

[0211] 指纹传感器1214用于采集用户的指纹,由处理器1201根据指纹传感器1214采集到的指纹识别用户的身份,或者,由指纹传感器1214根据采集到的指纹识别用户的身份。在识别出用户的身份为可信身份时,由处理器1201授权该用户执行相关的敏感操作,该敏感操作包括解锁屏幕、查看加密信息、下载软件、支付及更改设置等。指纹传感器1214可以被设置在电子设备1200的正面、背面或侧面。当电子设备1200上设置有物理按键或厂商Logo时,指纹传感器1214可以与物理按键或厂商Logo集成在一起。

[0212] 光学传感器1215用于采集环境光强度。在一个实施例中,处理器1201可以根据光学传感器1215采集的环境光强度,控制显示屏1205的显示亮度。具体地,当环境光强度较高时,调高显示屏1205的显示亮度;当环境光强度较低时,调低显示屏1205的显示亮度。在另一个实施例中,处理器1201还可以根据光学传感器1215采集的环境光强度,动态调整摄像头组件1206的拍摄参数。

[0213] 接近传感器1216,也称距离传感器,通常设置在电子设备1200的前面板。接近传感器1216用于采集用户与电子设备1200的正面之间的距离。在一个实施例中,当接近传感器1216检测到用户与电子设备1200的正面之间的距离逐渐变小时,由处理器1201控制显示屏1205从亮屏状态切换为息屏状态;当接近传感器1216检测到用户与电子设备1200的正面之间的距离逐渐变大时,由处理器1201控制显示屏1205从息屏状态切换为亮屏状态。

[0214] 本领域技术人员可以理解,图12中示出的结构并不构成对电子设备1200的限定,可以包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件布置。

[0215] 本申请还提供一种计算机可读存储介质,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现上述方法实施例提供的词权重生成方法。

[0216] 本申请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方法实施例提供的词权重生成方法。

[0217] 上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0218] 在一个实施例中,上述视频特征向量、文本特征向量和词语特征向量可存储在数据共享系统的节点中。参见图13所示的数据共享系统,数据共享系统1300是指用于进行节点与节点之间数据共享的系统,该数据共享系统中可以包括多个节点1301,多个节点1301可以是指数据共享系统中各个客户端。每个节点1301在进行正常工作可以接收到输入信息,并基于接收到的输入信息维护该数据共享系统内的共享数据。为了保证数据共享系统内的信息互通,数据共享系统中的每个节点之间可以存在信息连接,节点之间可以通过上

述信息连接进行信息传输。例如,当数据共享系统中的任意节点接收到输入信息时,数据共享系统中的其他节点便根据共识算法获取该输入信息,将该输入信息作为共享数据中的数据进行存储,使得数据共享系统中全部节点上存储的数据均一致。

[0219] 对于数据共享系统中的每个节点,均具有与其对应的节点标识,而且数据共享系统中的每个节点均可以存储有数据共享系统中其他节点的节点标识,以便后续根据其他节点的节点标识,将生成的区块广播至数据共享系统中的其他节点。每个节点中可维护一个如下表所示的节点标识列表,将节点名称和节点标识对应存储至该节点标识列表中。其中,节点标识可为IP (Internet Protocol,网络之间互联的协议) 地址以及其他任一种能够用于标识该节点的信息,下表中仅以IP地址为例进行说明。

节点名称	节点标识
节点1	117.114.151.174
节点2	117.116.189.145
...	...
节点N	119.123.789.258

[0221] 数据共享系统中的每个节点均存储一条相同的区块链。区块链由多个区块组成,参见图14,区块链由多个区块组成,创始块中包括区块头和区块主体,区块头中存储有输入信息特征值、版本号、时间戳和难度值,区块主体中存储有输入信息;创始块的下一区块以创始块为父区块,下一区块中同样包括区块头和区块主体,区块头中存储有当前区块的输入信息特征值、父区块的区块头特征值、版本号、时间戳和难度值,并以此类推,使得区块链中每个区块中存储的区块数据均与父区块中存储的区块数据存在关联,保证了区块中输入信息的安全性。

[0222] 在生成区块链中的各个区块时,参见图15,区块链所在的节点在接收到输入信息时,对输入信息进行校验,完成校验后,将输入信息存储至内存池中,并更新其用于记录输入信息的哈希树;之后,将更新时间戳更新为接收到输入信息的时间,并尝试不同的随机数,多次进行特征值计算,使得计算得到的特征值可以满足下述公式:

$$[0223] \quad \text{SHA256}(\text{SHA256}(\text{version}+\text{prev_hash}+\text{merkle_root}+\text{ntime}+\text{nbits}+\text{x})) < \text{TARGET}$$

[0224] 其中,SHA256为计算特征值所用的特征值算法;version(版本号)为区块链中相关区块协议的版本信息;prev_hash为当前区块的父区块的区块头特征值;merkle_root为输入信息的特征值;ntime为更新时间戳的更新时间;nbits为当前难度,在一段时间内为定值,并在超出固定时间段后再次进行确定;x为随机数;TARGET为特征值阈值,该特征值阈值可以根据nbits确定得到。

[0225] 这样,当计算得到满足上述公式的随机数时,便可将信息对应存储,生成区块头和区块主体,得到当前区块。随后,区块链所在节点根据数据共享系统中其他节点的节点标识,将新生成的区块分别发送给其所在的数据共享系统中的其他节点,由其他节点对新生成的区块进行校验,并在完成校验后将新生成的区块添加至其存储的区块链中。

[0226] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0227] 以上所述仅为本申请的可选实施例,并不用以限制本申请,凡在本申请的精神和

原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

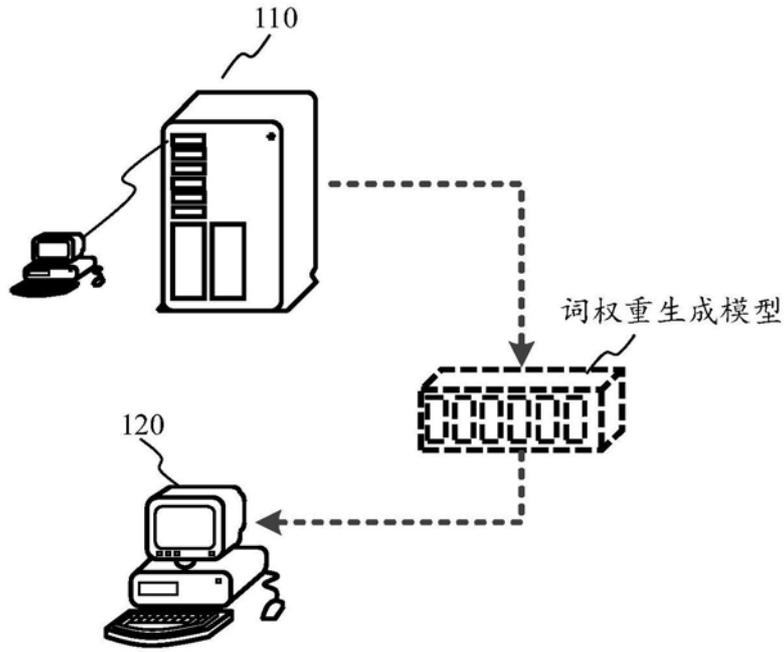


图1

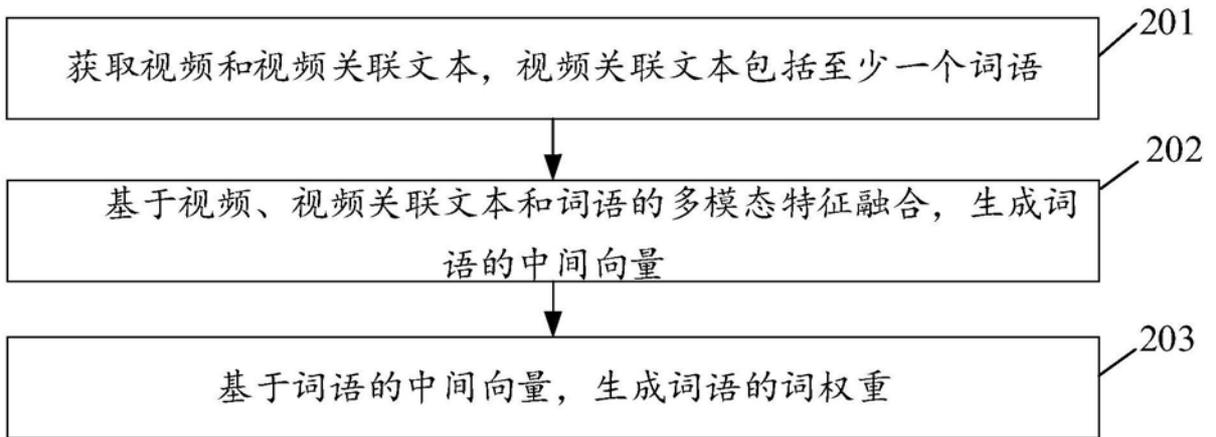


图2

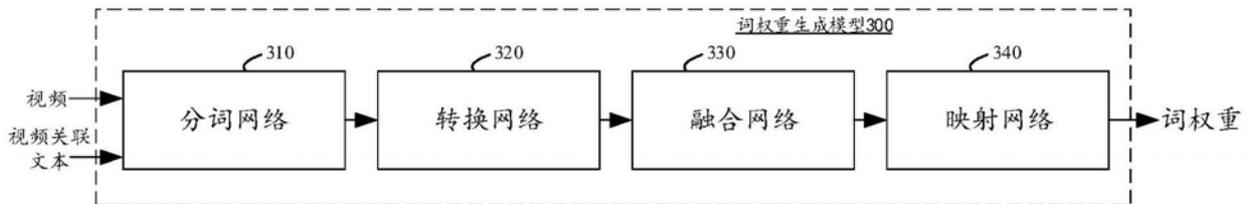


图3

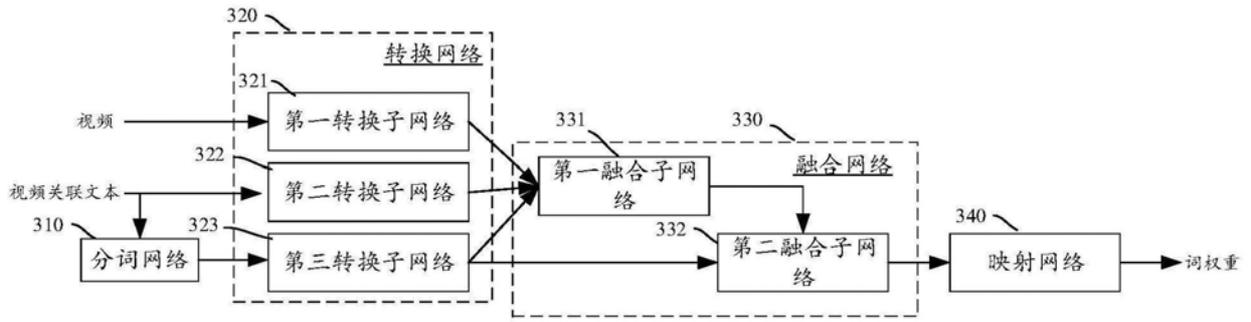


图4

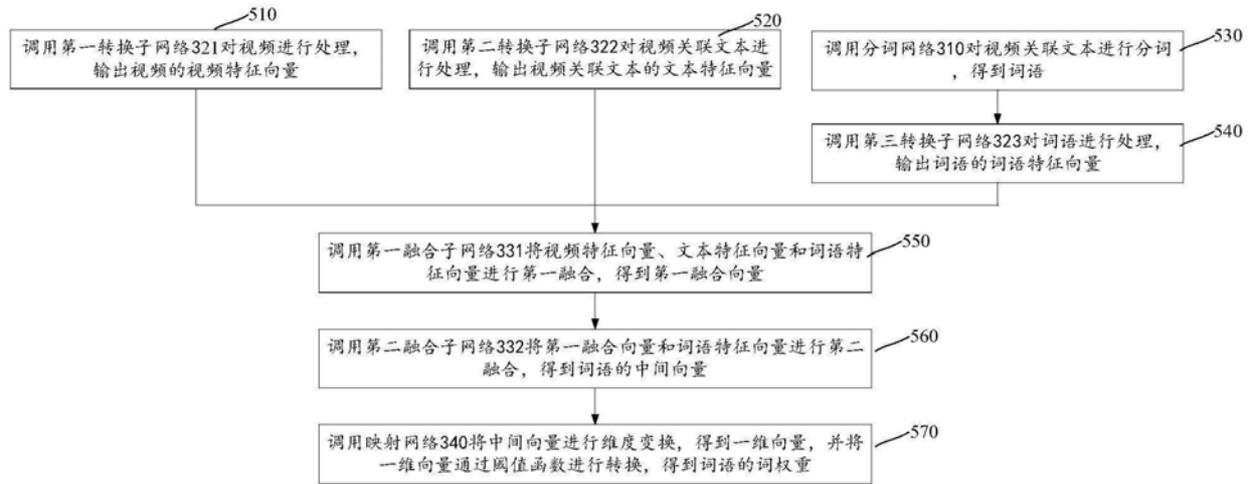


图5

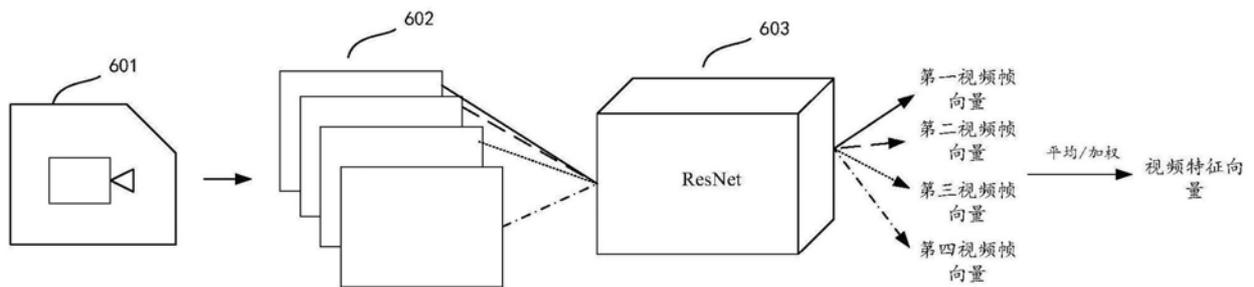


图6

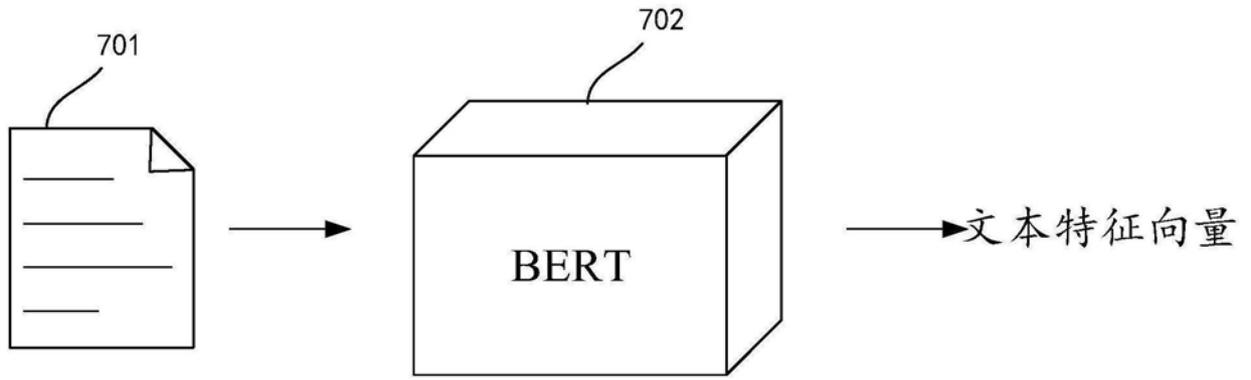


图7

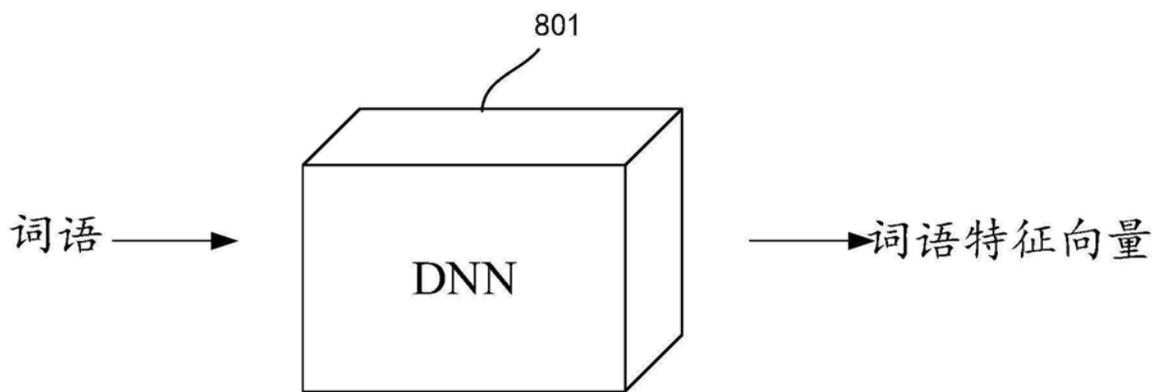


图8

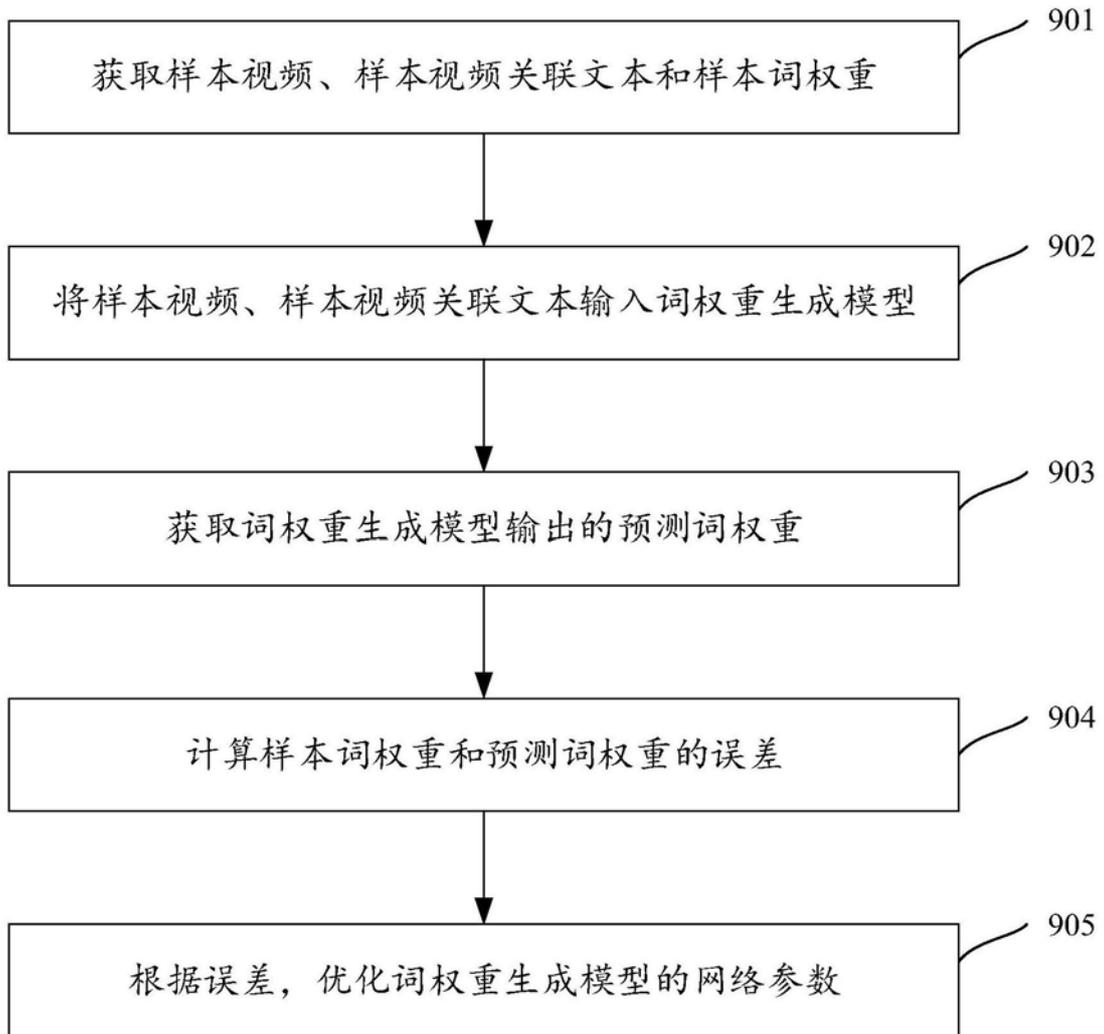


图9

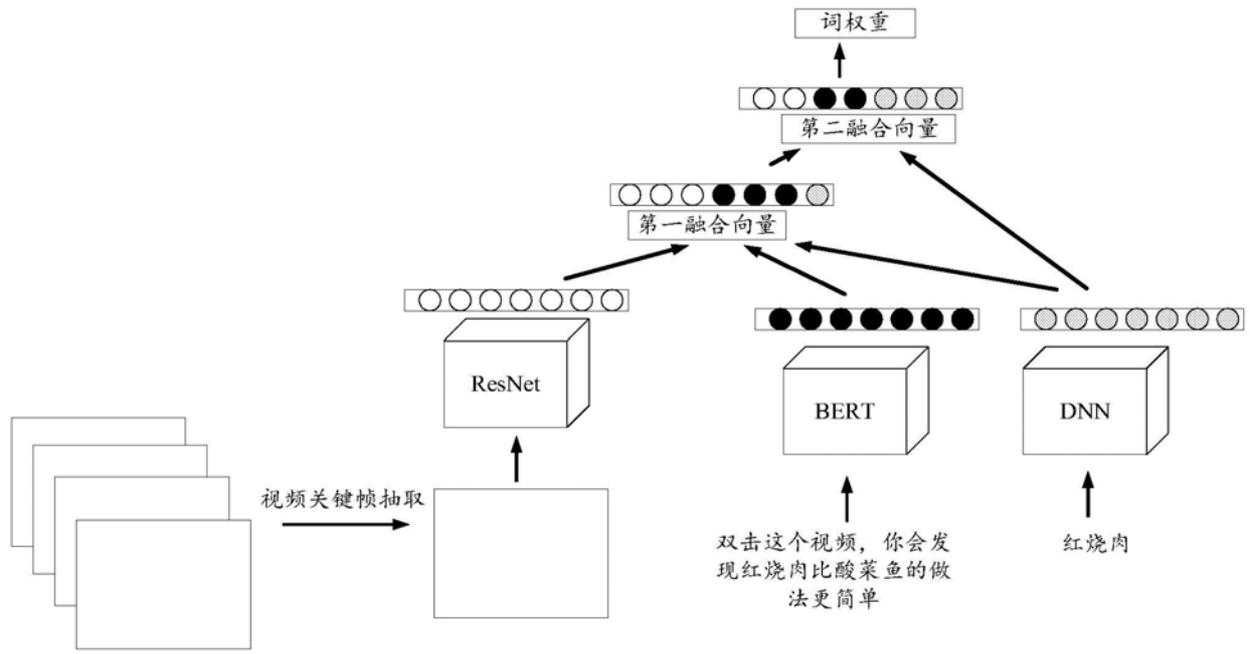


图10

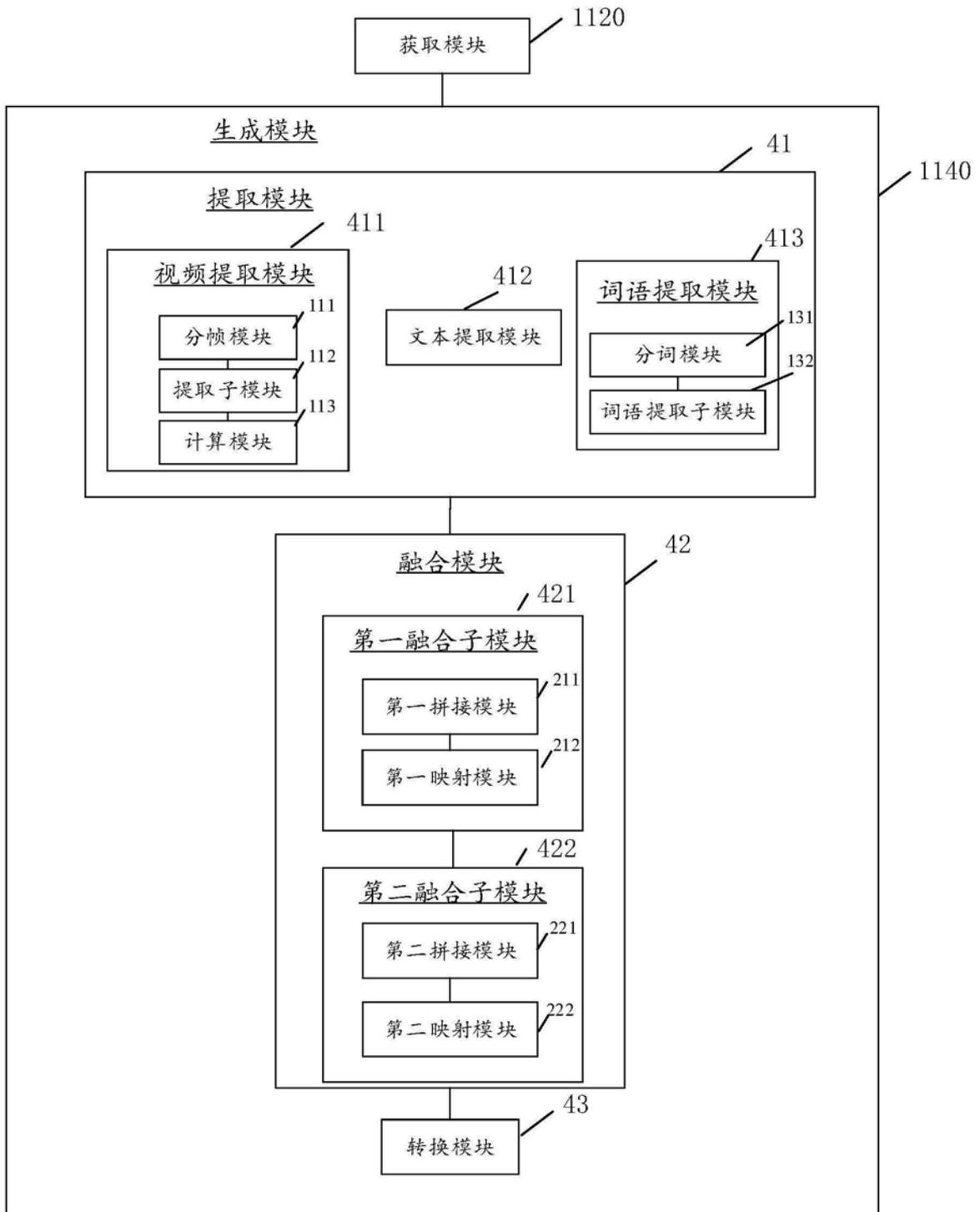


图11

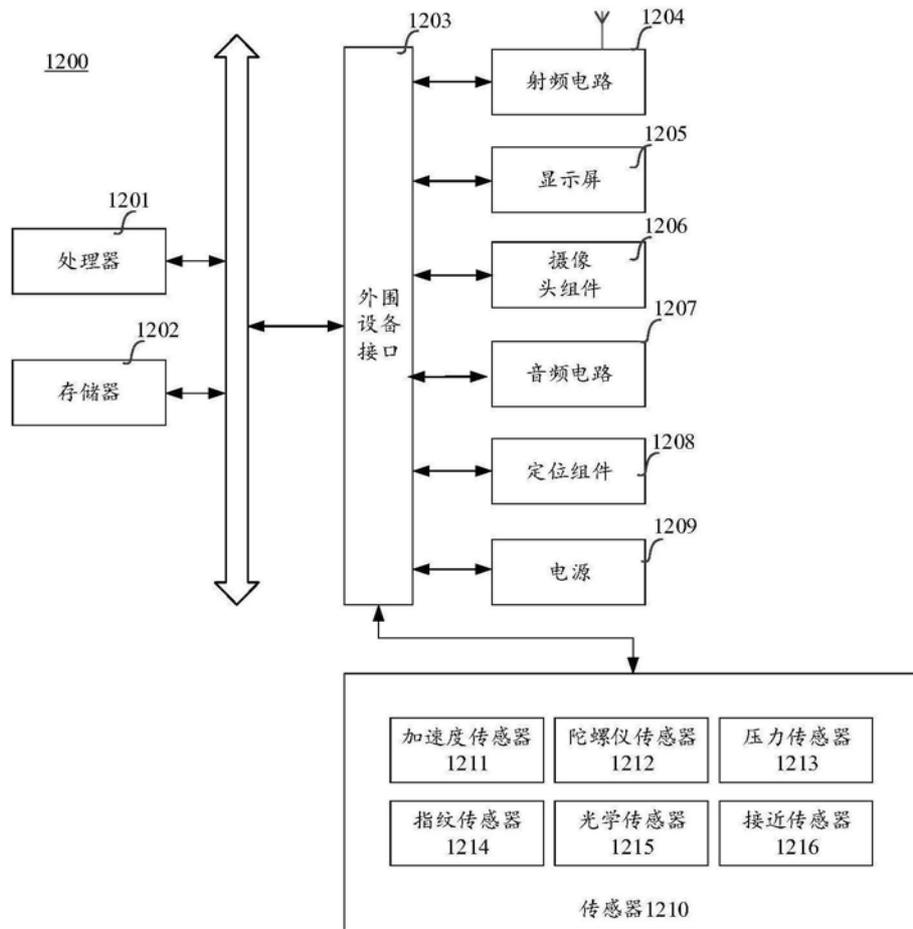


图12

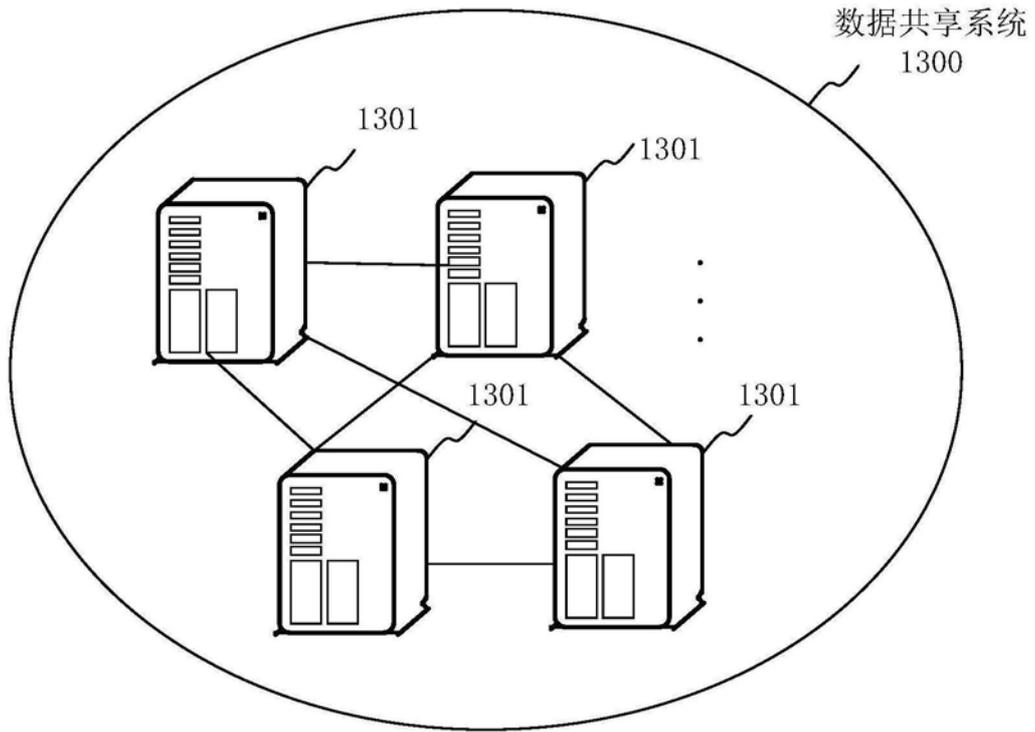


图13

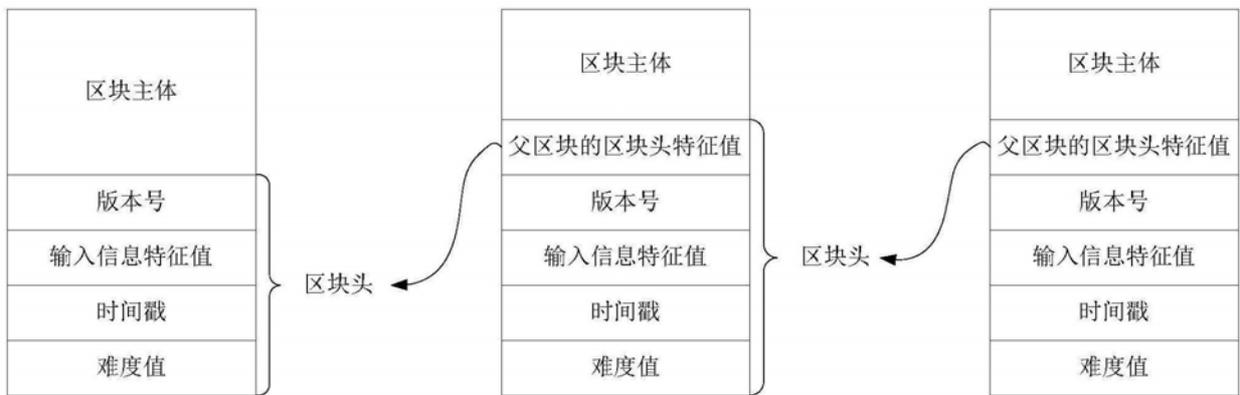


图14

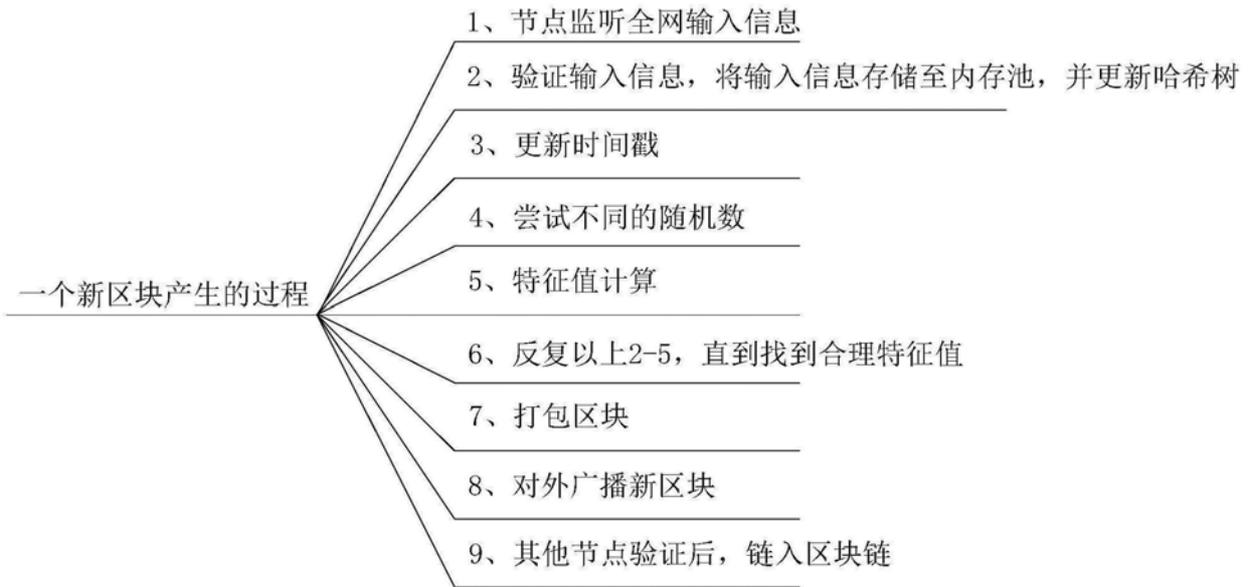


图15