(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113360711 A
(43)申请公布日 2021.09.07

(21)申请号 202110731643.3

(22)申请日 2021.06.29

(71)申请人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号百度大厦2层

(72)发明人 曲福 金志鹏 杨羿 陈晓冬 贺翔

(74)专利代理机构 北京市通商律师事务所 11951
代理人 巩靖

(51)Int.Cl.
G06F 16/78(2019.01)
G06F 16/75(2019.01)
G06F 40/30(2020.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

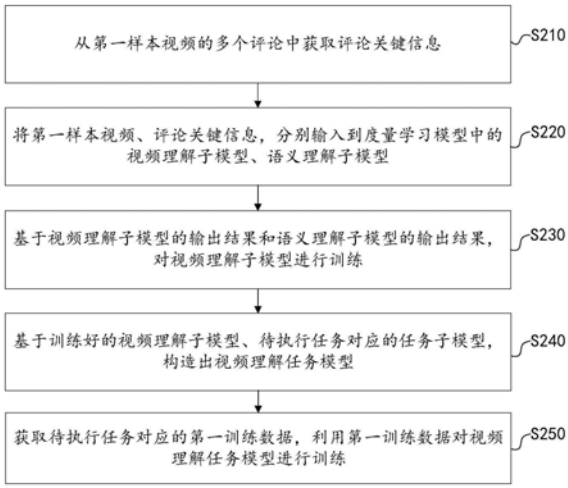
权利要求书3页 说明书10页 附图5页

(54)发明名称

视频理解任务的模型训练和执行方法、装置、设备及介质

(57)摘要

本公开提供了一种视频理解任务的模型训练和执行方法、装置、设备及介质，涉及人工智能领域，尤其涉及视频理解的领域。具体实现方案为：从第一样本视频的多个评论中获取评论关键信息；将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型；基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练；基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型；获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练。该方法提升了训练数据的获取效率，并且可以确保视频理解子模型可以较准确地对视频的内容进行理解。



CN 113360711 A

1.一种视频理解任务模型的训练方法,包括:

从第一样本视频的多个评论中获取评论关键信息;

将所述第一样本视频、所述所述评论关键信息,分别输入到度量学习模型中的视频理解子模型、语义理解子模型;

基于所述视频理解子模型的输出结果和所述语义理解子模型的输出结果,对所述视频理解子模型进行训练;

基于训练好的所述视频理解子模型、待执行任务对应的任务子模型,构造出视频理解任务模型;

获取所述待执行任务对应的第一训练数据,利用所述第一训练数据对所述视频理解任务模型进行训练。

2.根据权利要求1所述的方法,其中,所述从第一样本视频的多个评论中获取评论关键信息,包括:

获取所述第一样本视频的多个评论,从所述多个评论中确定出字数超过第一预设字数的多个有效评论;

从所述第一样本视频的多个有效评论中获取评论关键信息。

3.根据权利要求1所述的方法,其中,在所述从第一样本视频的多个评论中获取评论关键信息之前,还包括:

获取多个候选视频,并确定每个所述候选视频的评论数量;

将评论数量超过第一预设数量的所述候选视频,确定为第一样本视频。

4.根据权利要求1所述的方法,其中,所述基于所述视频理解子模型的输出结果和所述语义理解子模型的输出结果,对所述视频理解子模型进行训练,包括:

利用所述视频理解子模型输出第一表征向量,利用所述语义理解子模型输出第二表征向量;

确定出所述第一表征向量和所述第二表征向量的相似度,基于所述相似度调整所述视频理解子模型和所述语义理解子模型的参数。

5.根据权利要求1至4中任一项所述的方法,其中,所述视频理解子模型的结构类型包括:基于帧特征的Transformer结构、基于目标底层特征的Transformer结构、三维卷积神经网络结构。

6.根据权利要求1至4中任一项所述的方法,其中,所述语义理解子模型的结构类型至少包括基于文本关键词的Transformer结构。

7.根据权利要求1至4中任一项所述的方法,其中,所述待执行任务至少包括视频分类任务、视频搜索任务、视频推荐任务和广告匹配任务;

所述任务子模型至少包括对应于所述视频分类任务的分类子模型、对应于所述视频搜索任务的搜索子模型、对应于所述视频推荐任务的推荐子模型和对应于所述广告匹配任务的匹配子模型。

8.一种针对视频的任务执行方法,包括:

获取待执行任务的任务数据,将所述任务数据输入到根据权利要求1至7任一项所述训练方法得到的视频理解任务模型;

利用所述视频理解任务模型输出任务结果。

9.一种视频理解任务模型的训练装置,包括:

评论信息获取模块,用于从第一样本视频的多个评论中获取评论关键信息;

评论信息输入模块,用于将所述第一样本视频、所述所述评论关键信息,分别输入到度量学习模型中的视频理解子模型、语义理解子模型;

第一模型训练模块,用于基于所述视频理解子模型的输出结果和所述语义理解子模型的输出结果,对所述视频理解子模型进行训练;

模型构造模块,用于基于训练好的所述视频理解子模型、待执行任务对应的任务子模型,构造出视频理解任务模型;

第二模型训练模块,用于获取所述待执行任务对应的第一训练数据,利用所述第一训练数据对所述视频理解任务模型进行训练。

10.根据权利要求9所述的装置,其中,所述评论信息获取模块在用于从第一样本视频的多个评论中获取评论关键信息时,具体用于:

获取所述第一样本视频的多个评论,从所述多个评论中确定出字数超过第一预设字数的多个有效评论;

从所述第一样本视频的多个有效评论中获取评论关键信息。

11.根据权利要求9所述的装置,还包括样本筛选模块,所述样本筛选模块用于:

获取多个候选视频,并确定每个所述候选视频的评论数量;

将评论数量超过第一预设数量的所述候选视频,确定为第一样本视频。

12.根据权利要求9所述的装置,其中,所述第一模型训练模块在用于基于所述视频理解子模型的输出结果和所述语义理解子模型的输出结果,对所述视频理解子模型进行训练时,具体用于:

利用所述视频理解子模型输出第一表征向量,利用所述语义理解子模型输出第二表征向量;

确定出所述第一表征向量和所述第二表征向量的相似度,基于所述相似度调整所述视频理解子模型和所述语义理解子模型的参数。

13.一种针对视频的任务执行装置,包括:

任务输入模块,用于获取待执行任务的任务数据,将所述任务数据输入到根据权利要求1至7任一项所述训练方法得到的视频理解任务模型;

任务执行模块,用于利用所述视频理解任务模型输出任务结果。

14.一种电子设备,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能够执行权利要求1-7中任一项所述的方法。

15.一种电子设备,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能够执行权利要求8所述的方法。

16.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-7中任一项所述的方法。

17.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求8所述的方法。

18.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-7中任一项所述的方法。

19.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求8所述的方法。

视频理解任务的模型训练和执行方法、装置、设备及介质

技术领域

[0001] 本公开涉及人工智能领域，尤其涉及视频理解的领域，可以应用在视频分类、视频搜索、视频推荐和广告匹配等场景中。

背景技术

[0002] 视频理解模型是一种能够理解视频内容的人工智能，在现有的针对视频理解模型的训练过程中，获取训练数据需要耗费大量的人工成本，而且，所获取到的训练数据所包含的信息通常较少，导致训练数据中的信息较为片面，不利于提升视频理解模型的。

发明内容

[0003] 本公开提供了一种视频理解任务的模型训练和执行方法、装置、设备及介质。

[0004] 根据本公开的第一方面，提供了一种视频理解任务模型的训练方法，包括：

[0005] 从第一样本视频的多个评论中获取评论关键信息；

[0006] 将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型；

[0007] 基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练；

[0008] 基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型；

[0009] 获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练。

[0010] 根据本公开的第二方面，提供了一种针对视频的任务执行方法，包括：

[0011] 获取待执行任务的任务数据，将任务数据输入到根据权利要求1至7任一项训练方法得到的视频理解任务模型；

[0012] 利用视频理解任务模型输出任务结果。

[0013] 根据本公开的第三方面，提供了一种视频理解任务模型的训练装置，包括：

[0014] 评论信息获取模块，用于从第一样本视频的多个评论中获取评论关键信息；

[0015] 评论信息输入模块，用于将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型；

[0016] 第一模型训练模块，用于基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练；

[0017] 模型构造模块，用于基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型；

[0018] 第二模型训练模块，用于获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练。

[0019] 根据本公开的第四方面，提供了一种针对视频的任务执行装置，包括：

[0020] 任务输入模块，用于获取待执行任务的任务数据，将任务数据输入到根据本公开的第一方面训练方法得到的视频理解任务模型；

[0021] 任务执行模块，用于利用视频理解任务模型输出任务结果。

[0022] 根据本公开的第五方面，提供了一种电子设备，包括：

[0023] 至少一个处理器；以及

[0024] 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能够执行上述的视频理解任务模型的训练方法。

[0025] 根据本公开的第六方面，提供了一种电子设备，包括：

[0026] 至少一个处理器；以及

[0027] 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能够执行上述的针对视频的任务执行方法。

[0028] 根据本公开的第七方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，其中，所述计算机指令用于使所述计算机执行上述的视频理解任务模型的训练方法。

[0029] 根据本公开的第八方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，其中，所述计算机指令用于使所述计算机执行上述的针对视频的任务执行方法。

[0030] 根据本公开的第九方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现上述的视频理解任务模型的训练方法。

[0031] 根据本公开的第十方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现上述的针对视频的任务执行方法。

[0032] 应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

[0033] 本公开提供的技术方案带来的有益效果是：

[0034] 本公开实施提供的方案，本方案自动获取视频和视频的评论作为训练数据对度量学习模型进行训练，提升了训练数据的获取效率；基于评论可以丰富训练数据的数据量。

附图说明

[0035] 附图用于更好地理解本方案，不构成对本公开的限定。其中：

[0036] 图1示出了本公开实施例提供的一种度量学习模型的示意性结构图；

[0037] 图2示出了本公开实施例提供的一种视频理解任务模型的训练方法的流程示意图；

[0038] 图3示出了本公开实施例提供的另一种视频理解任务模型的训练方法的流程示意图；

[0039] 图4示出了本公开实施例提供的一种针对视频的任务执行方法的流程示意图；

[0040] 图5示出了本公开实施例提供的一种视频理解任务模型的训练装置的结构示意图之一；

[0041] 图6示出了本公开实施例提供的一种视频理解任务模型的训练装置的结构示意图之二；

[0042] 图7示出了本公开实施例提供的一种针对视频的任务执行装置的结构示意图；

[0043] 图8示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。

具体实施方式

[0044] 以下结合附图对本公开的示范性实施例做出说明，其中包括本公开实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本公开的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

[0045] 本公开实施例提供的视频理解任务的模型训练和执行方法、装置、设备及介质，旨在解决现有技术的如上技术问题中的至少一个。

[0046] 图1示出了本公开实施例提供的一种度量学习模型的示意性结构图。在此需要说明的是，在数学中，一个度量(或距离函数)是一个定义集合中元素之间距离的函数，度量学习(Metric Learning)的基本原理是根据不同的任务来自主学习出针对某个特定任务的度量距离函数。如图1所示，度量学习模型包括第一输入层、第一表示层、第二输入层、第二表示层和匹配层。其中，第一输入层同于输入视频，第一表示层用于理解视频的内容，第二输入层用于输入视频的评论信息，第二表示层用于理解视频的评论信息的内容，匹配层用于对第一表示层的输出结果进行预设处理(计算相似度)，以便根据处理结果调整度量学习模型的参数。

[0047] 本公开实施例中的度量学习模型可以是双塔模型，其中，第一输入层和第一表示层为视频侧塔，第二输入层和第二表示层为评论侧塔，为了便于表述，将视频侧塔定义为视频理解子模型，将评论侧塔定义为语义理解子模型。

[0048] 可选地，视频理解子模型的结构类型可以是基于帧特征的Transformer结构、基于目标底层特征的Transformer结构、三维卷积神经网络结构之中的任一项，当然，视频理解子模型的结构还可以为其他类型，此处不一一列举。

[0049] 可选地，语义理解子模型的结构类型可以是基于文本关键词的Transformer结构，当然，语义理解子模型的结构还可以为其他类型，此处不一一列举。

[0050] 图2示出了本公开实施例提供的一种视频理解任务模型的训练方法的流程示意图，如图2所示，该方法主要可以包括以下步骤：

[0051] S210：从第一样本视频的多个评论中获取评论关键信息。

[0052] 在本公开实施例中，可以对第一样本视频的评论进行筛选，仅保留信息量较多的有效评论作为训练数据，确保训练结果的准确性、提升训练效率。可选地，可以通过评论所包含的字数确定评论是否为有效评论，具体地，可以获取第一样本视频的多个评论，从多个评论中确定出字数超过第一预设字数的多个有效评论，从第一样本视频的多个有效评论中获取评论关键信息。

[0053] 在本公开实施例中，在步骤S210之前，还可以对视频进行评论进行筛选，仅保留有效的视频作为第一样本视频。可选地，可以通过视频的评论数量来确定该视频否为有效的视频，具体地，可以获取多个候选视频，并确定每个候选视频的评论数量；将评论数量超过第一预设数量的候选视频，确定为第一样本视频。

[0054] S220：将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子

模型、语义理解子模型。

[0055] 在本公开实施例中，将第一样本视频到度量学习模型中的视频理解子模型，具体来说，可以将第一样本视频通过度量学习模型的第一输入模型输入到第一表示层。

[0056] 在本公开实施例中，将评论关键信息输入到度量学习模型中的语义理解子模型，具体来说，可以将评论关键信息通过度量学习模型的第二输入模型输入到第二表示层。

[0057] S230：基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练。

[0058] 在本公开实施例中，视频理解子模型的输出结果和语义理解子模型的输出结果均为用于表征视频内容的表征向量，可以通过两个表征向量的比较结果来调整度量学习模型的参数。

[0059] 可选地，可以利用视频理解子模型输出第一表征向量，利用语义理解子模型输出第二表征向量，确定出第一表征向量和第二表征向量的相似度，基于相似度调整视频理解子模型和语义理解子模型的参数。可以理解，两个表征向量的相似度越高，表示视频理解子模型所理解的视频内容月准确，调整视频理解子模型和语义理解子模型的参数目的是使得两个表征向量的相似度可以达到期望的相似度值。视频理解子模型和语义理解子模型所输出的表征向量的相似度，可以比较客观准确地体现视频理解子模型和语义理解子模型对视频理解的差异程度，基于相似度能够针对性地调整模型的参数，以便较快地使模型到达预期的效果。

[0060] S240：基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型。

[0061] 可以理解，训练好的视频理解子模型可以较准确地对视频进行理解，基于该视频理解子模型可以执行一些实际的待执行任务。在本公开实施例中，待执行任务至少包括视频分类任务、视频搜索任务、视频推荐任务和广告匹配任务，但不限于此。

[0062] 每种待执行任务需要一个任务子模型与视频理解子模型构成一个完整的视频理解子模型，利用视频理解子模型来完成待执行任务。可选地，任务子模型至少包括对应于视频分类任务的分类子模型、对应于视频搜索任务的搜索子模型、对应于视频推荐任务的推荐子模型和对应于广告匹配任务的匹配子模型。以分类子模型为例，视频理解子模型和分类子模型所构成的视频理解任务模型，可以用于对视频进行分类。

[0063] S250：获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练。

[0064] 以视频理解任务模型是由视频理解子模型和分类子模型所构成的为例，该视频理解任务模型可以用于对视频进行分类。第一训练数据可以是视频，为了便于理解和表述，可以将该视频定义为第二样本视频，将第二样本视频输入到视频理解任务模型中的视频理解子模型，利用视频理解子模型输出表征第二样本视频的内容的表征向量，之后将表征向量输入到视频理解任务模型中的分类子模型，利用分类子模型确定出第二样本视频的类型结果，基于类型结果的准确度来调整视频理解任务模型的参数，以便视频理解任务模型所确定的视频类型的准确率达到期望的准确率。

[0065] 本公开实施提供的视频理解任务模型的训练方法，预先构建了包含视频理解子模型和语义理解子模型的度量学习模型，自动获取视频和视频的评论作为训练数据对度量学

习模型进行训练，提升了训练数据的获取效率；由于评论中可以包含与视频相关的大量信息，极大地丰富了训练数据的数据量，确保训练数据的全面性和客观性，使得视频理解子模型可以较准确地对视频的内容进行理解。此外，将有训练好视频理解子模型应用到基于视频理解的下游任务，有助于提升下游任务的效果。

[0066] 图3示出了本公开实施例提供的另一种视频理解任务模型的训练方法的流程示意图，如图3所示，该方法主要可以包括以下步骤：

[0067] S310：获取多个候选视频，并确定每个候选视频的评论数量。

[0068] 本公开实施例可以收集大量视频及其对应的评论并存储在数据库，可以从数据库拉取多个候选视频，构建包含如下样本的数据集：

[0069] 样本1：Id1\t content1\t content2\t……contentn；

[0070] 样本2：Id2\t content1\t content2\t……contentn；

[0071] 以样本1为例，Id1是第一个候选视频的身份信息，“t content1”是第一个候选视频的第一条评论。在上述数据集中，针对每个候选视频，统计出该候选视频的评论数量。

[0072] S320：将评论数量超过第一预设数量的候选视频，确定为第一样本视频。

[0073] 可以理解，第一预设数量的数值可以根据实际的设计需要而定，例如，第一预设数量可以是200，则可以将评论数量200的候选视频确定为第一样本视频。

[0074] S330：获取第一样本视频的多个评论，从多个评论中确定出字数超过第一预设字数的多个有效评论。

[0075] 可以理解，第一预设字数的数值可以根据实际的设计需要而定，例如，第一预设字数可以是15，则可以加工第一样本视频的多个评论中字数超过15的评论确定为有效评论。

[0076] S340：从第一样本视频的多个有效评论中获取评论关键信息。

[0077] 在本公开实施例中，可以从第一样本视频的多个有效评论，按照预设的提取规则提取关键词，将提取到的关键词作为评论关键信息。

[0078] S350：将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型。

[0079] 在本公开实施例中，将第一样本视频到度量学习模型中的视频理解子模型，具体来说，可以将第一样本视频通过度量学习模型的第一输入模型输入到第一表示层。

[0080] 在本公开实施例中，将评论关键信息输入到度量学习模型中的语义理解子模型，具体来说，可以将评论关键信息通过度量学习模型的第二输入模型输入到第二表示层。

[0081] S360：利用视频理解子模型输出第一表征向量，利用语义理解子模型输出第二表征向量。

[0082] 可选地，视频理解子模型和语义理解子模型可以将第一表征向量和第二表征向量输入到度量学习模型的匹配层中，以便匹配层对第一表征向量和第二表征向量进行预设处理。

[0083] S370：确定出第一表征向量和第二表征向量的相似度，基于相似度调整视频理解子模型和语义理解子模型的参数。

[0084] 可选地，可以利用度量学习模型的匹配层确定出第一表征向量和第二表征向量的相似度，该相似度可以是余弦相似度。可以理解，两个表征向量的相似度越高，表示视频理解子模型所理解的视频内容月准确，调整视频理解子模型和语义理解子模型的参数目的是

使得两个表征向量的相似度可以达到期望的相似度值。

[0085] S380:基于训练好的视频理解子模型、待执行任务对应的任务子模型,构造出视频理解任务模型。

[0086] 在本公开实施例中,步骤S380的具体描述可以参照步骤S240中的描述,此处不再赘述。

[0087] S390:获取待执行任务对应的第一训练数据,利用第一训练数据对视频理解任务模型进行训练。

[0088] 在本公开实施例中,步骤S390的具体描述可以参照步骤S250中的描述,此处不再赘述。

[0089] 图4示出了本公开实施例提供的一种针对视频的任务执行方法的流程示意图,如图4所示,该方法主要可以包括以下步骤:

[0090] S410:获取待执行任务的任务数据,将任务数据输入到根据视频理解任务模型的训练方法得到的视频理解任务模型。

[0091] S420:利用视频理解任务模型输出任务结果。

[0092] 在本公开实施例中,待执行任务至少包括视频分类任务、视频搜索任务、视频推荐任务和广告匹配任务,但不限于此。

[0093] 每种待执行任务需要一个任务子模型与视频理解子模型构成一个完整的视频理解子模型,利用视频理解子模型来完成待执行任务。可选地,任务子模型至少包括对应于视频分类任务的分类子模型、对应于视频搜索任务的搜索子模型、对应于视频推荐任务的推荐子模型和对应于广告匹配任务的匹配子模型。

[0094] 以分类子模型为例,视频理解子模型和分类子模型所构成的视频理解任务模型,可以用于对视频进行分类。待执行任务的任务数据可以是视频,为了便于理解和表述,将该视频称为待分类视频,可以将待分类视频输入到视频理解任务模型中的视频理解子模型,利用视频理解子模型输出表征待分类视频的内容的表征向量,之后将表征向量输入到视频理解任务模型中的分类子模型,利用分类子模型确定出待分类视频的类型结果。

[0095] 基于与上述的视频理解任务模型的训练方法相同的原理,图5示出了本公开实施例提供的一种视频理解任务模型的训练装置的结构示意图之一,图6示出了本公开实施例提供的一种视频理解任务模型的训练装置的结构示意图之二。如图5所示,视频理解任务模型的训练装置500包括评论信息获取模块510、评论信息输入模块520、第一模型训练模块530、模型构造模块540和第二模型训练模块550。

[0096] 评论信息获取模块510用于从第一样本视频的多个评论中获取评论关键信息。

[0097] 评论信息输入模块520用于将第一样本视频、评论关键信息,分别输入到度量学习模型中的视频理解子模型、语义理解子模型。

[0098] 第一模型训练模块530用于基于视频理解子模型的输出结果和语义理解子模型的输出结果对视频理解子模型进行训练。

[0099] 模型构造模块540用于基于训练好的视频理解子模型、待执行任务对应的任务子模型,构造出视频理解任务模型。

[0100] 第二模型训练模块550,用于获取待执行任务对应的第一训练数据,利用第一训练数据对视频理解任务模型进行训练。

[0101] 本公开实施提供的视频理解任务模型的训练装置，预先构建了包含视频理解子模型和语义理解子模型的度量学习模型，自动获取视频和视频的评论作为训练数据对度量学习模型进行训练，提升了训练数据的获取效率；由于评论中可以包含与视频相关的大量信息，极大地丰富了训练数据的数据量，确保训练数据的全面性和客观性，使得视频理解子模型可以较准确地对视频的内容进行理解。此外，将有训练好视频理解子模型应用到基于视频理解的下游任务，有助于提升下游任务的效果。

[0102] 在本公开实施例中，评论信息获取模块510在用于从第一样本视频的多个评论中获取评论关键信息时，具体用于：

[0103] 获取第一样本视频的多个评论，从多个评论中确定出字数超过第一预设字数的多个有效评论；

[0104] 从第一样本视频的多个有效评论中获取评论关键信息。

[0105] 在本公开实施例中，如图6所示，视频理解任务模型的训练装置500还包括样本筛选模块560，样本筛选模块560用于：

[0106] 获取多个候选视频，并确定每个候选视频的评论数量；

[0107] 将评论数量超过第一预设数量的候选视频，确定为第一样本视频。

[0108] 在本公开实施例中，第一模型训练模块530在用于基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练时，具体用于：

[0109] 利用视频理解子模型输出第一表征向量，利用语义理解子模型输出第二表征向量；

[0110] 确定出第一表征向量和第二表征向量的相似度，基于相似度调整视频理解子模型和语义理解子模型的参数。

[0111] 在本公开实施例中，视频理解子模型的结构类型包括：基于帧特征的Transformer结构、基于目标底层特征的Transformer结构、三维卷积神经网络结构。

[0112] 在本公开实施例中，语义理解子模型的结构类型至少包括基于文本关键词的Transformer结构。

[0113] 在本公开实施例中，待执行任务至少包括视频分类任务、视频搜索任务、视频推荐任务和广告匹配任务；

[0114] 任务子模型至少包括对应于视频分类任务的分类子模型、对应于视频搜索任务的搜索子模型、对应于视频推荐任务的推荐子模型和对应于广告匹配任务的匹配子模型。

[0115] 可以理解的是，本公开实施例中的视频理解任务模型的训练装置500的上述各模块具有实现上述的视频理解任务模型的训练方法相应步骤的功能。该功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。该硬件或软件包括一个或多个与上述功能相对应的模块。上述模块可以是软件和/或硬件，上述各模块可以单独实现，也可以多个模块集成实现。对于上述视频理解任务模型的训练装置500的各模块的功能描述具体可以参见上述的内容推荐方法的对应描述，在此不再赘述。

[0116] 基于与上述的针对视频的任务执行方法相同的原理，图7示出了本公开实施例提供的一种针对视频的任务执行装置的结构示意图。如图7所示，针对视频的任务执行装置700包括任务输入模块710和任务执行模块720。

[0117] 任务输入模块710用于获取待执行任务的任务数据，将任务数据输入到根据视频

理解任务模型的训练方法得到的视频理解任务模型。

[0118] 任务执行模块720用于利用视频理解任务模型输出任务结果。

[0119] 可以理解的是，本公开实施例中的针对视频的任务执行装置的上述各模块具有实现上述的针对视频的任务执行方法相应步骤的功能。该功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。该硬件或软件包括一个或多个与上述功能相对应的模块。上述模块可以是软件和/或硬件，上述各模块可以单独实现，也可以多个模块集成实现。对于上述针对视频的任务执行装置的各模块的功能描述具体可以参见上述的模型训练方法的对应描述，在此不再赘述。

[0120] 本公开的技术方案中，所涉及的用户个人信息的获取，存储和应用等，均符合相关法律法规的规定，且不违背公序良俗。

[0121] 根据本公开的实施例，本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0122] 图8示出了可以用来实施本公开的实施例的示例电子设备的示意性框图，可以理解，电子设备可以用来实施本公开的实施例视频理解任务模型的训练方法和针对视频的任务执行方法中的至少一个。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0123] 如图8所示，设备800包括计算单元801，其可以根据存储在只读存储器(ROM)802中的计算机程序或者从存储单元808加载到随机访问存储器(RAM)803中的计算机程序，来执行各种适当的动作和处理。在RAM 803中，还可存储设备800操作所需的各种程序和数据。计算单元801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0124] 设备800中的多个部件连接至I/O接口805，包括：输入单元806，例如键盘、鼠标等；输出单元807，例如各种类型的显示器、扬声器等；存储单元808，例如磁盘、光盘等；以及通信单元809，例如网卡、调制解调器、无线通信收发机等。通信单元809允许设备800通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0125] 计算单元801可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元801的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元801执行上文所描述的各个方法和处理，例如视频理解任务模型的训练方法和针对视频的任务执行方法中的至少一个。例如，在一些实施例中，视频理解任务模型的训练方法和针对视频的任务执行方法中的至少一个可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元808。在一些实施例中，计算机程序的部分或者全部可以经由ROM 802和/或通信单元809而被载入和/或安装到设备800上。当计算机程序加载到RAM 803并由计算单元801执行时，可以执行上文描述的视频理解任务模型的训练方法的一个或多个步骤，或者可以执行上文描述的针对视频的任务执行方法的一个或多个步骤，在其他实施例中，计算单元801可以通过其他任何适当的方

式(例如,借助于固件)而被配置为执行视频理解任务模型的训练方法和针对视频的任务执行方法之中的至少一个。

[0126] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0127] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0128] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0129] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0130] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0131] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器,也可以为分布式系统的

服务器，或者是结合了区块链的服务器。

[0132] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本公开公开的技术方案所期望的结果，本文在此不进行限制。

[0133] 上述具体实施方式，并不构成对本公开保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等，均应包含在本公开保护范围之内。

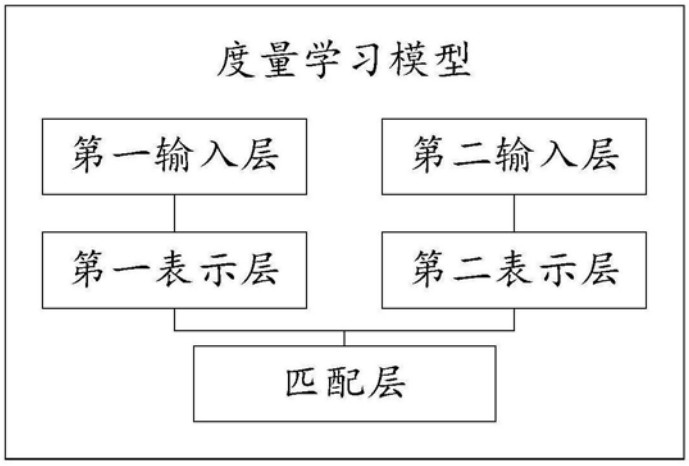

度量学习模型
第一输入层
第二输入层
第一表示层
第二表示层
匹配层

图1

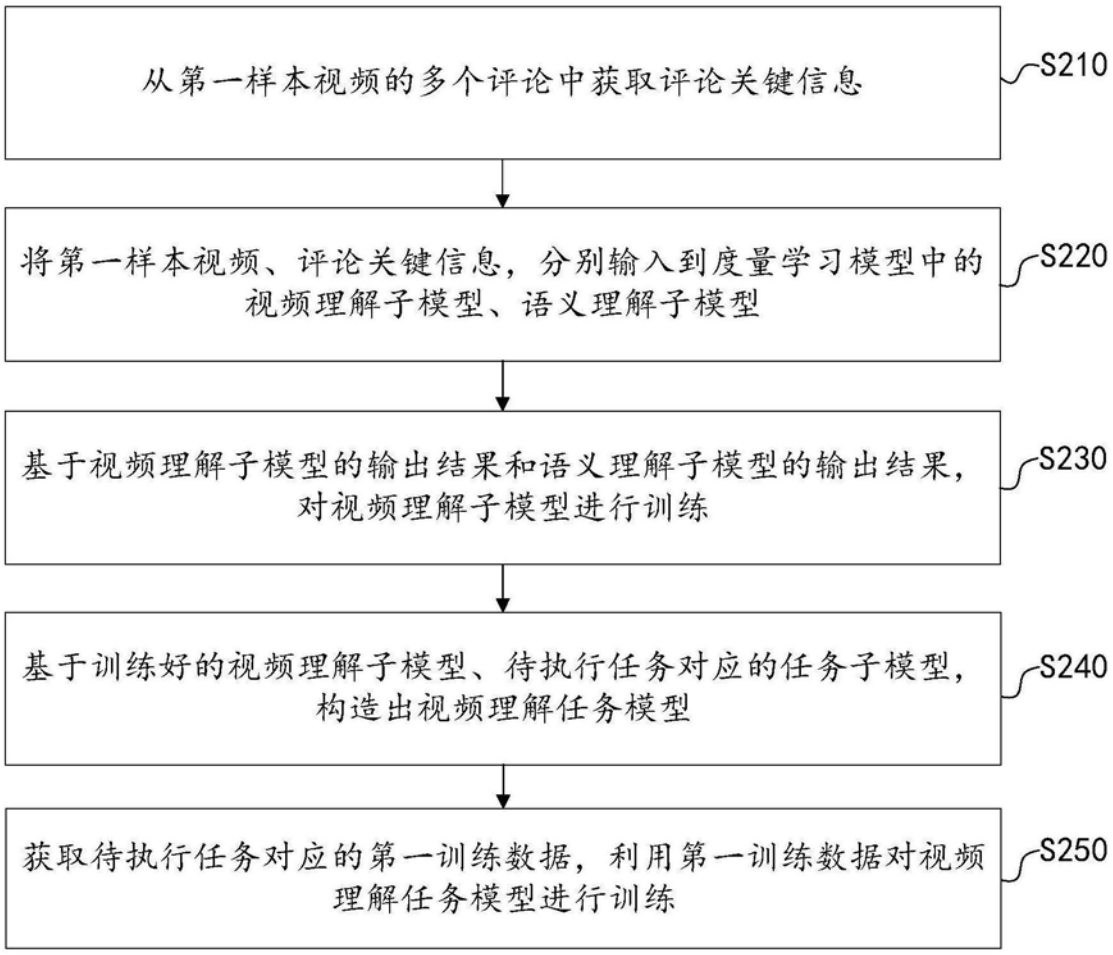

从第一样本视频的多个评论中获取评论关键信息
S210
将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型
S220
基于视频理解子模型的输出结果和语义理解子模型的输出结果，对视频理解子模型进行训练
S230
基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型
S240
获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练
S250

图2

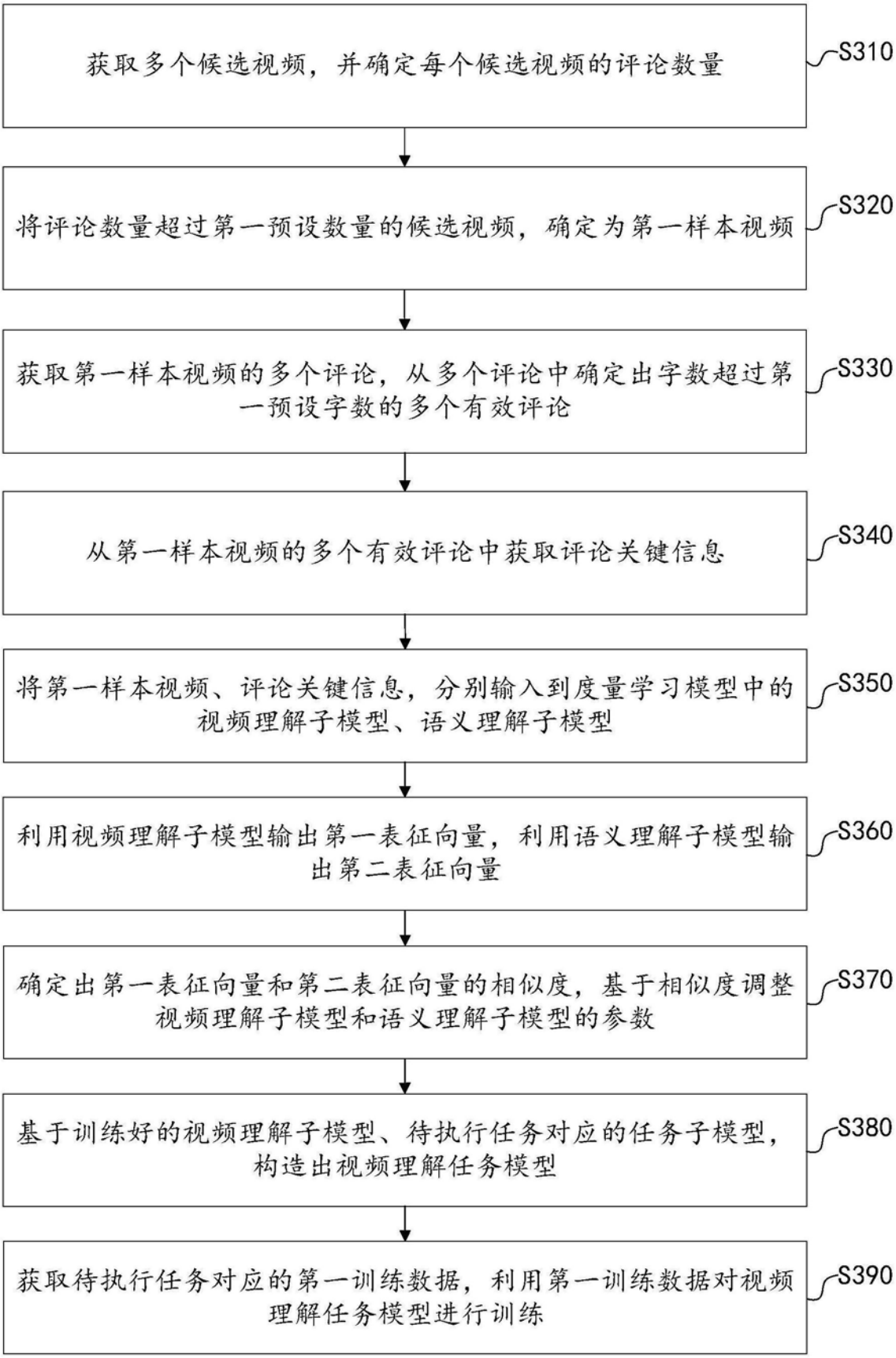

获取多个候选视频，并确定每个候选视频的评论数量
S310
将评论数量超过第一预设数量的候选视频，确定为第一样本视频
S320
获取第一样本视频的多个评论，从多个评论中确定出字数超过第一预设字数的多个有效评论
S330
从第一样本视频的多个有效评论中获取评论关键信息
S340
将第一样本视频、评论关键信息，分别输入到度量学习模型中的视频理解子模型、语义理解子模型
S350
利用视频理解子模型输出第一表征向量，利用语义理解子模型输出第二表征向量
S360
确定出第一表征向量和第二表征向量的相似度，基于相似度调整视频理解子模型和语义理解子模型的参数
S370
基于训练好的视频理解子模型、待执行任务对应的任务子模型，构造出视频理解任务模型
S380
获取待执行任务对应的第一训练数据，利用第一训练数据对视频理解任务模型进行训练
S390

图3

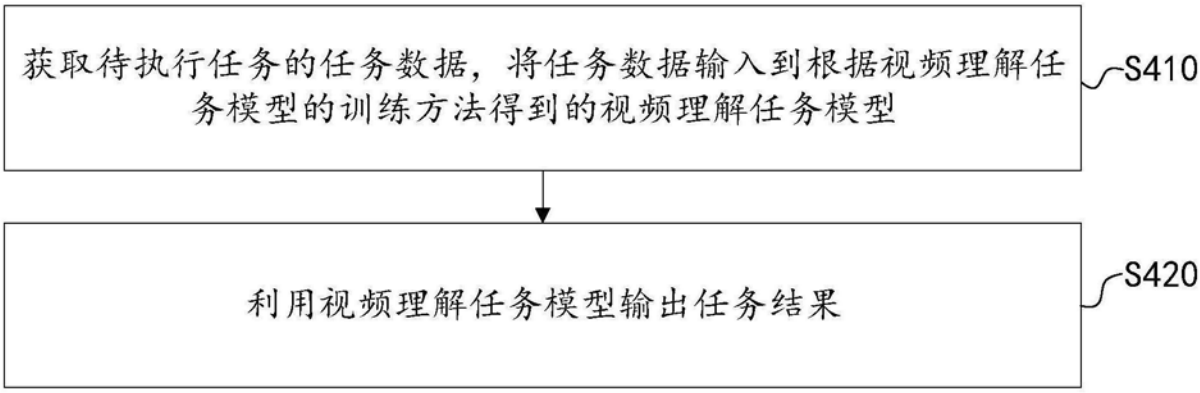

获取待执行任务的任务数据，将任务数据输入到根据视频理解任务模型的训练方法得到的视频理解任务模型
S410
利用视频理解任务模型输出任务结果
S420

图4

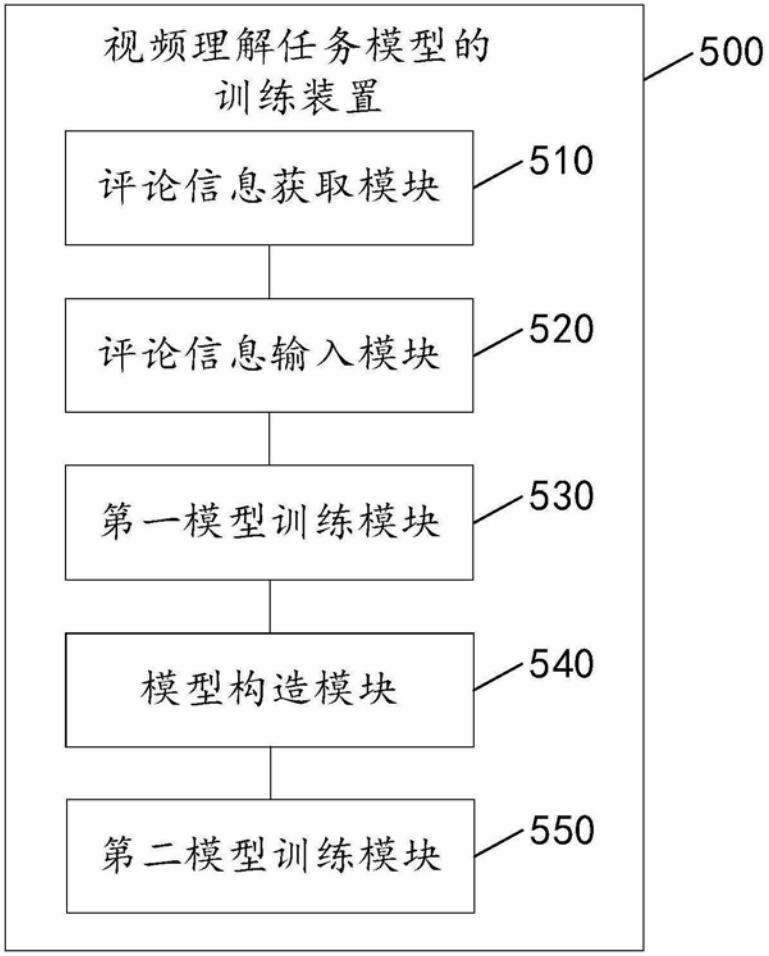

视频理解任务模型的训练装置
500
评论信息获取模块
510
评论信息输入模块
520
第一模型训练模块
530
模型构造模块
540
第二模型训练模块
550

图5

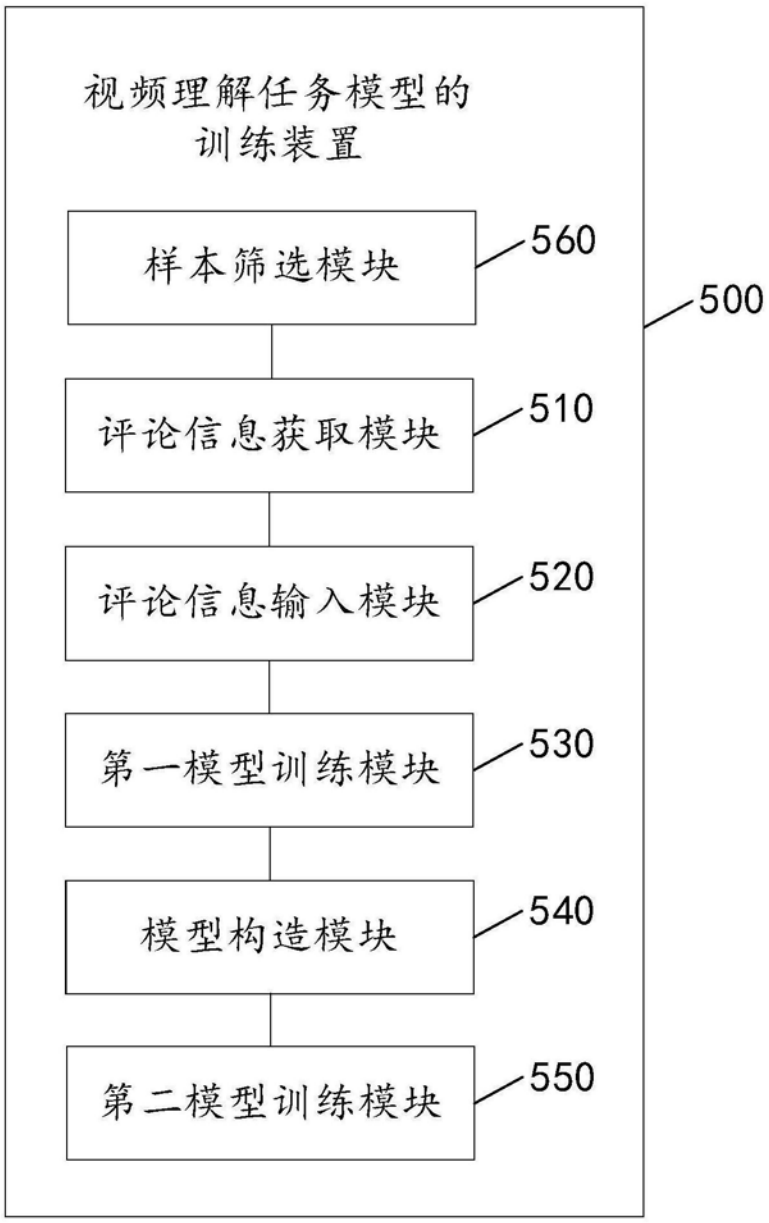

视频理解任务模型的
训练装置
样本筛选模块
560
500
评论信息获取模块
510
评论信息输入模块
520
第一模型训练模块
530
模型构造模块
540
第二模型训练模块
550

图6

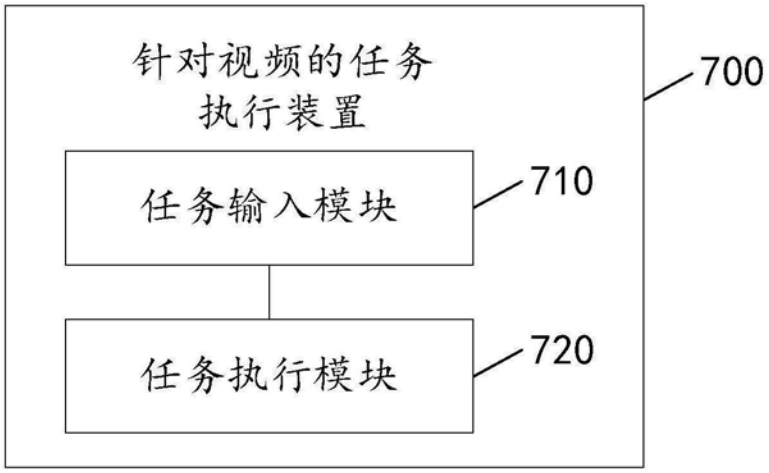

针对视频的任务
执行装置
700
任务输入模块
710
任务执行模块
720

图7

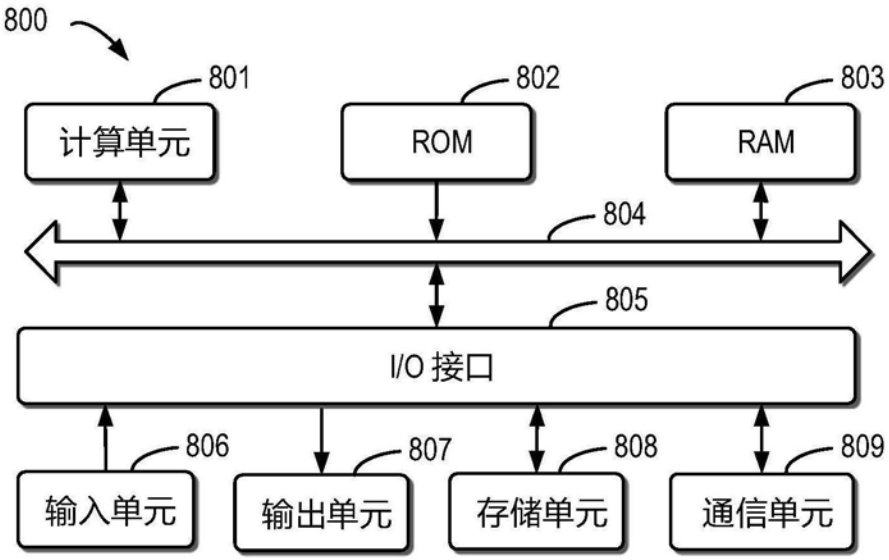

800
801
计算单元
802
ROM
803
RAM
804
805
I/O 接口
806
输入单元
807
输出单元
808
存储单元
809
通信单元

图8