

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 11 月 28 日 (28.11.2024)



(10) 国际公布号
WO 2024/239783 A1

- (51) 国际专利分类号:
G06F 16/21 (2019.01) G06F 16/2457 (2019.01)
G06F 16/23 (2019.01)
- (21) 国际申请号: PCT/CN2024/083332
- (22) 国际申请日: 2024 年 3 月 22 日 (22.03.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310572043.6 2023年5月19日 (19.05.2023) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳

市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

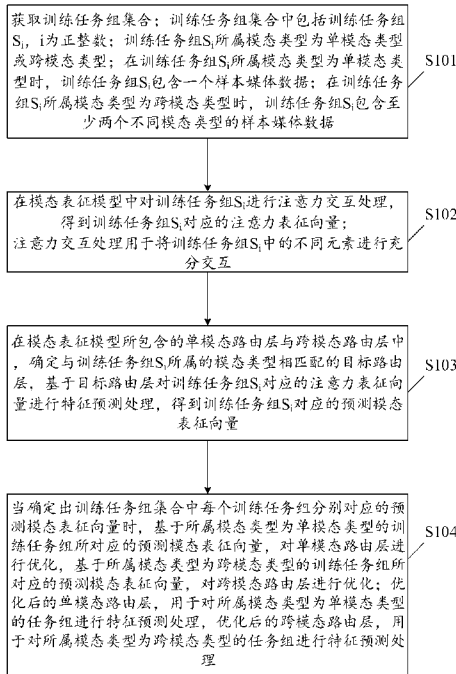
(72) 发明人: 刘焯 (LIU, Ye); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路 80 号汇华商贸大厦 1508 室, Guangdong 510070 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,

(54) Title: DATA PROCESSING METHOD AND APPARATUS, DEVICE, AND READABLE STORAGE MEDIUM

(54) 发明名称: 一种数据处理方法、装置、设备以及可读存储介质



- S101 Acquire a set of training task groups, the set of training task groups comprising a training task group S_i , wherein i is a positive integer; the modal type of the training task group S_i is a unimodal type or a cross-modal type; when the modal type of the training task group S_i is the unimodal type, the training task group S_i comprises a piece of sample media data; and when the modal type of the training task group S_i is the cross-modal type, the training task group S_i comprises at least two different modal types of sample media data
- S102 Perform attention interaction processing on the training task group S_i in a modal representation model to obtain an attention representation vector corresponding to the training task group S_i , wherein the attention interaction processing is used for performing full interaction on different elements in the training task group S_i
- S103 In a unimodal routing layer and a cross-modal routing layer comprised in the modal representation model, determine a target routing layer matched with the modal type of the training task group S_i , and on the basis of the target routing layer, perform feature prediction processing on the attention representation vector corresponding to the training task group S_i , to obtain a prediction modal representation vector corresponding to the training task group S_i
- S104 When prediction modal representation vectors respectively corresponding to training task groups in the set of training task groups are determined, optimize the unimodal routing layer on the basis of prediction modal representation vectors corresponding to training task groups of which the modal type is the unimodal type, and optimize the cross-modal routing layer on the basis of prediction modal representation vectors corresponding to training task groups of which the modal type is the cross-modal type, wherein the optimized unimodal routing layer is used for performing feature prediction processing on task groups of which the modal type is the unimodal type, and the optimized cross-modal routing layer is used for performing feature prediction processing on task groups of which the modal type is the cross-modal type

图 2

(57) Abstract: The present application discloses a data processing method and apparatus, a device, and a readable storage medium. The method comprises: acquiring a set of training task groups, the set of training task groups comprising a training task group S_i ; performing attention interaction processing on the training task group S_i to obtain an attention representation vector corresponding to the training task group S_i ; determining a target routing layer matched with the modal type of the training task group S_i , and performing feature prediction processing

[见续页]



LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN,
MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,
PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,
SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

on the attention representation vector on the basis of the target routing layer to obtain a prediction modal representation vector corresponding to the training task group S_i ; and optimizing a unimodal routing layer on the basis of prediction modal representation vectors corresponding to training task groups of which the modal type is a unimodal type, and optimizing a cross-modal routing layer on the basis of prediction modal representation vectors corresponding to training task groups of which the modal type is a cross-modal type. By adopting the present application, the universality of a modal representation model can be improved, and the accuracy of a modal representation result is improved.

(57) 摘要: 本申请公开了一种数据处理方法、装置、设备以及可读存储介质, 方法包括: 获取训练任务组集合; 训练任务组集合中包括训练任务组 S_i ; 对训练任务组 S_i 进行注意力交互处理, 得到训练任务组 S_i 对应的注意力表征向量; 确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层, 基于目标路由层对注意力表征向量进行特征预测处理, 得到训练任务组 S_i 对应的预测模态表征向量; 基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量, 对单模态路由层进行优化, 基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量, 对跨模态路由层进行优化。采用本申请, 可以提升模态表征模型的通用性, 提升模态表征结果的准确性。

一种数据处理方法、装置、设备以及可读存储介质

本申请要求于 2023 年 05 月 19 日提交中国专利局、申请号为 2023105720436、申请名称为“一种数据处理方法、装置、设备以及可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，尤其涉及一种数据处理方法、装置、设备以及可读存储介质。

背景技术

随着多媒体技术的快速发展，媒体数据（如图片、文本、视频、音频等等）得到海量生产，而随着海量媒体数据的生成，媒体数据的理解任务（如视频理解任务）就变得尤为重要，媒体数据的理解任务能够为媒体数据提供众多丰富且多样的媒体标签（如舞蹈标签、唱歌标签、竞赛标签、游戏标签等），通过媒体数据的媒体标签可以便于对媒体数据进行多种媒体处理，例如，可以便于对媒体数据进行媒体检索、分类、归档、进行媒体推荐、媒体二次编辑等处理。

在媒体检索任务中，通过媒体标签可以很便利地检索出相关的媒体片段，媒体数据的文本标签、标题信息等，媒体检索对于媒体推荐、媒体加工都有着非常重要的意义。而随着媒体检索需求越来越大，对媒体检索的能力要求也越来越高，例如，在输入一个模态类型为文本类型的媒体数据后，对于检索结果的需求不仅限于是模态类型为文本类型的媒体数据，还需求检索到模态类型为非文本类型（如视频类型、音频类型、图像类型等）的媒体数据。即，在媒体数据的检索业务中，基于多模态的检索性能要求越来越大。

然而，在相关技术中，对于媒体数据多模态的理解任务，会由一个统一的模型来进行理解推理，具体的，针对各个模态会提取每个模态的特征，之后，再将所有模态的特征多模态信息进行融合，最终可以得到多模态信息。但是，相关技术在进行模型的训练时，为了提升多模态信息的准确性，会剔除单模态的个性信息，并提取不同模态之间的共性信息，由此会严重影响模型对于单模态数据的理解准确性，从而严重影响单模态数据在后续任务（如检索任务）中的处理结果的准确性。

发明内容

本申请实施例提供一种数据处理方法、装置、设备以及可读存储介质，可以在模态表征业务中，提升模态表征模型的通用性，提升模态表征结果的准确性。

本申请实施例一方面提供了一种数据处理方法，包括：

获取训练任务组集合；训练任务组集合中包括训练任务组 S_i ， i 为正整数；所述训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在所述训练任务组 S_i 所属模态类型为所述单模态类型时，所述训练任务组 S_i 包含一个样本媒体数据；在所述训练任务组 S_i 所属模态类型为所述跨模态类型时，所述训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量；注意力交互处理用于将训练任务组 S_i 中的不同元素进行充分交互；

在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层，基于目标路由层对训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到训练任务组 S_i 对应的预测模态表征向量；

当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为单

模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化，基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为跨模态类型的任务组进行特征预测处理。

本申请实施例一方面提供了一种数据处理装置，包括：

集合获取模块，用于获取训练任务组集合；训练任务组集合中包括训练任务组 S_i ， i 为正整数；所述训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在所述训练任务组 S_i 所属模态类型为所述单模态类型时，所述训练任务组 S_i 包含一个样本媒体数据；在所述训练任务组 S_i 所属模态类型为所述跨模态类型时，所述训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

注意力处理模块，用于在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量；注意力交互处理用于将训练任务组 S_i 中的不同元素进行充分交互；

路由层确定模块，用于在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

预测模块，用于基于目标路由层对训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到训练任务组 S_i 对应的预测模态表征向量；

第一优化模块，用于当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化；

第二优化模块，用于基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为跨模态类型的任务组进行特征预测处理。

在一个实施例中，集合获取模块获取训练任务组集合的具体实现方式，包括：

获取 N 个样本媒体数据； N 为正整数， N 个媒体数据中的一个样本媒体数据所属的模态类型，为第一模态类型与第二模态类型中的任意一种；第一模态类型与第二模态类型不同；

基于 N 个样本媒体数据中，每个样本媒体数据分别所属的模态类型，对 N 个样本媒体数据进行任务组构建处理，得到训练任务组集合。

在一个实施例中，集合获取模块基于 N 个样本媒体数据中，每个样本媒体数据分别所属的模态类型，对 N 个样本媒体数据进行任务组构建处理，得到训练任务组集合的具体实现方式，包括：

将 N 个样本媒体数据的第一样本媒体数据，与用于表征第一模态类型的第一标识进行组合，得到所属模态类型为第一模态类型的第一训练任务组；第一样本媒体数据是指 N 个样本媒体数据中，所属模态类型为第一模态类型的任一样本媒体数据；

将 N 个样本媒体数据中的第二样本媒体数据，与用于表征第二模态类型的第二标识进行组合，得到所属模态类型为第二模态类型的第二训练任务组；第二样本媒体数据是指 N 个样本媒体数据中，所属模态类型为第二模态类型的任一样本媒体数据；

基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道，对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理，得到所属模态类型为跨模态类型的训练任务组；

将第一训练任务组与第二训练任务组，均确定为所属模态类型为单模态类型的训练任务组，将所属模态类型为所述跨模态类型的训练任务组，与所属模态类型为所述单模态类型的训练任务组所组成的集合，确定为训练任务组集合。

在一个实施例中，第一样本媒体数据与第二样本媒体数据的数量均为一个或多个；一个或多个第一样本媒体数据中包括第一样本媒体数据 M_j ， j 为正整数；所属模态类型为跨模态类型的训练任务组，包含第一样本媒体数据 M_j 对应的训练任务组；

集合获取模块基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道，对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理，得到所属模态类型为跨模态类型的训练任务组的具体实现方式，包括：

将第一样本媒体数据 M_j 对应的媒体来源渠道，确定为目标媒体来源渠道；

将一个或多个第二样本媒体数据中，媒体来源渠道为目标媒体来源渠道的第二样本媒体数据，确定为第一样本媒体数据 M_j 对应的关联样本媒体数据；

将第一样本媒体数据、关联样本媒体数据、第一标识以及第二标识进行组合，得到第一样本媒体数据 M_j 对应的训练任务组。

在一个实施例中，注意力处理模块在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量的具体实现方式，包括：

在模态表征模型中，通过特征提取网络层对训练任务组 S_i 所包含的样本媒体数据进行特征提取处理，得到训练任务组 S_i 对应的媒体特征；

通过模态表征模型中的多头自注意力网络层，对训练任务组 S_i 对应的媒体特征进行多头自注意力处理，得到训练任务组 S_i 对应的注意力表征向量。

在一个实施例中，模态表征模型中的多头自注意力网络层包括 Q 个自注意力子网络层； Q 为正整数； Q 个自注意力子网络层包括自注意力子网络层 V_k ， k 为正整数；

注意力处理模块通过模态表征模型中的多头自注意力网络层，对训练任务组 S_i 对应的媒体特征进行多头自注意力处理，得到训练任务组 S_i 对应的注意力表征向量的具体实现方式，包括：

获取自注意力子网络层 V_k 所包含的注意力参数矩阵，将自注意力子网络层 V_k 所包含的注意力参数矩阵，与训练任务组 S_i 对应的媒体特征进行运算处理，得到注意力参数矩阵对应的线性变换矩阵；

通过自注意力子网络层 V_k 中的全连接组件，对线性变换矩阵进行特征整合处理，得到自注意力子网络层 V_k 对应的注意力表征子向量；

当确定出 Q 个自注意力子网络层分别对应的注意力表征子向量时，将 Q 个注意力表征子向量进行融合处理，得到训练任务组 S_i 对应的注意力表征向量。

在一个实施例中，单模态类型包含第一模态类型与第二模态类型；单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层；

路由层确定模块在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层的具体实现方式，包括：

若训练任务组 S_i 所属的模态类型为跨模态类型，则将模态表征模型中所包含的跨模态路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若训练任务组 S_i 所属的模态类型为第一模态类型，则将第一模态类型对应的第一模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若训练任务组 S_i 所属的模态类型为第二模态类型，则将第二模态类型对应的第二模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层。

在一个实施例中，单模态类型包含第一模态类型与第二模态类型；单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层；

第一优化模块基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化的具体实现方式，包括：

将所属模态类型为第一模态类型的训练任务组，确定为第一模态训练任务组，将所属模态类型为第二模态类型的训练任务组，确定为第二模态训练任务组；

获取第一模态训练任务组对应的真实模态表征向量，以及第二模态训练任务组对应的真实模态表征向

量；

对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值；

对第二模态训练任务组对应的预测模态表征向量，与第二模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第二模态训练任务组对应的第二损失值；

基于第一损失值对第一模态类型对应的第一模态子路由层进行优化，基于第二损失值对第二模态类型对应的第二模态子路由层进行优化。

在一个实施例中，第一模态类型为文本模态类型，第一模态训练任务组为文本词序列；第一模态训练任务组对应的预测模态表征向量由文本词序列中，每个文本词分别对应的预测表征特征所组成；第一模态训练任务组对应的真实模态表征向量，由文本词序列中，每个文本词分别对应的真实表征特征所组成；

第一优化模块对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值的具体实现方式，包括：

在文本词序列所包含的文本词中获取遮掩文本词；遮掩文本词是指经过遮掩处理的文本词；

在第一模态训练任务组对应的预测模态表征向量中，获取遮掩文本词对应的预测表征特征，在第一模态训练任务组对应的真实模态表征向量中，获取遮掩文本词对应的真实表征特征；

确定遮掩文本词对应的预测表征特征，与遮掩文本词对应的真实表征特征之间的第一特征相似度；

将第一特征相似度确定为第一模态训练任务组对应的第一损失值。

在一个实施例中，训练任务组集合中，包含至少两个所属模态类型为跨模态类型的训练任务组；至少两个所属模态类型为跨模态类型的训练任务组，包含训练任务组 S_j ， j 为正整数；训练任务组 S_j 包括所属模态类型为第一模态类型的第一样本媒体数据，以及所属模态类型为第二模态类型的第二样本媒体数据；训练任务组 S_j 所对应的预测模态表征向量中，包括第一样本媒体数据对应的预测表征特征，以及第二样本媒体数据对应的预测表征特征；

第二优化模块基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化的具体实现方式，包括：

确定第一样本媒体数据对应的预测表征特征，与第二样本媒体数据对应的预测表征特征之间的第二特征相似度；

将每个所属模态类型为跨模态类型的训练任务组，均确定为跨模态训练任务组，确定训练任务组 S_j 与至少两个跨模态训练任务组中，剩余跨模态训练任务组之间的第三特征相似度；剩余跨模态训练任务组是指至少两个跨模态训练任务组中，除训练任务组 S_j 以外的跨模态训练任务组；

基于第二特征相似度与第三特征相似度，对跨模态路由层进行优化。

在一个实施例中，数据处理装置还可以包括：

任务组获取模块，用于获取用于描述待分类媒体数据的目标任务组；目标任务组包括至少两个不同模态类型的媒体数据；

交互处理模块，用于在模态表征模型中对目标任务组进行注意力交互处理，得到目标任务组对应的注意力表征向量；

特征处理模块，用于在优化后的跨模态路由层中，对目标任务组对应的注意力表征向量进行特征预测处理，得到目标任务组对应的预测模态表征向量；

类别识别模块，用于对目标任务组对应的预测模态表征向量进行类别识别处理，得到待分类媒体数据所属的媒体类别；待分类媒体数据所属的媒体类别用于对待分类媒体数据进行归档处理。

本申请实施例一方面提供了一种计算机设备，包括：处理器和存储器；

存储器存储有计算机程序，计算机程序被处理器执行时，使得处理器执行本申请实施例中的方法。

本申请实施例一方面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程序，计算机程序包括程序指令，程序指令当被处理器执行时，执行本申请实施例中的方法。

本申请的一个方面，提供了一种计算机程序产品，该计算机程序产品包括计算机程序，该计算机程序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程序，处理器执行该计算机程序，使得该计算机设备执行本申请实施例一方面提供的方法。

在本申请实施例中，提供一种通用的模态表征模型，可以同时为所属模态类型为单模态类型的数据以及所属模态类型为跨模态类型的数据进行服务。具体的，本申请可以采用包含单模态类型与跨模态类型的训练任务组集合，对模态表征模型进行训练，其中，在模态表征模型中，包含有单模态路由层与跨模态路由层，单模态路由层可以对所属模态类型为单模态类型的数据（注意力表征向量）进行特征预测处理，跨模态路由层可以对所属模态类型为跨模态类型的数据（注意力表征向量）进行特征预测处理，而在训练优化时，单模态路由层可以基于所属模态类型为所述单模态类型的训练任务组所对应的预测模态表征向量来进行优化，跨模态路由层可以由所属模态类型为所述跨模态类型的训练任务组所对应的预测模态表征向量来进行优化。基于此，可以使得模态表征模型中的单模态路由层与跨模态路由层能够同时进行训练优化，且在训练优化时，是基于不同模态类型的数据来分别进行训练优化，无需损失任何一种模态类型的数据，可以同时保持单模态路由层与跨模态路由层的性能，由此可以提升模态表征模型对单模态类型与跨模态类型的通用性与灵活性。综上可知，本申请可以在模态表征模型中设置路由分发机制，将路由层划分为单模态路由层与跨模态路由层，不同的路由层可以对不同模态类型的数据进行处理，在训练时也由不同模态类型的数据分别进行训练优化，每个路由层只会保留当前模态类型的相关信息，由此无需舍弃任何一种模态类型的数据，也能够同时输出跨模态类型与单模态类型的表征向量，可以很好地提升各种模态类型的数据的表征结果的准确性。综上，本申请可以在模态表征业务中，提升模态表征模型的通用性，提升模态表征结果的准确性。

附图说明

- 图1是本申请实施例提供的一种数据处理系统的网络架构图；
- 图2是本申请实施例提供的一种数据处理方法的流程示意图；
- 图3是本申请实施例提供的一种 transformer 架构的架构示意图；
- 图4是本申请实施例提供的一种模态路由分发的架构示意图；
- 图5是本申请实施例提供的一种媒体数据检索的场景示意图；
- 图6是本申请实施例提供的一种对单模态路由层进行优化的流程示意图；
- 图7是本申请实施例提供的一种对跨模态路由层进行优化的流程示意图；
- 图8是本申请实施例提供的一种数据处理装置的结构示意图；
- 图9是本申请实施例提供的一种计算机设备的结构示意图。

具体实施方式

本申请涉及人工智能及相关概念，为便于理解，以下将对人工智能及相关概念进行简要阐述：

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。

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/深度学习等几大方向。

随着人工智能技术研究和进步，人工智能技术在多个领域展开研究和应用，例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等，相信随着技术的发展，人工智能技术将在更多的领域得到应用，并发挥越来越重要的价值。

本申请实施例提供的方案属于人工智能领域下属的机器学习(Machine Learning, ML)和自然语言处理(Nature Language processing, NLP)技术。

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。

自然语言处理(Nature Language processing, NLP)是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言，所以它与语言学的研究有着密切的联系。自然语言处理技术通常包括文本处理、语义理解、机器翻译、机器人问答、知识图谱等技术。

本申请实施例具体可通过自然语言处理技术，将媒体数据（如文本数据、视频数据）的相关文本数据（如标题数据、描述数据）进行语义分析处理，以得到该媒体数据的语义分析结果，从而能够更好地理解该媒体数据；同时，本申请实施例可以通过机器学习技术将相关模型（如后续提及的模态表征模型）进行训练优化，以提升模型的输出结果的准确性。

为便于理解，请参见图 1，图 1 是本申请实施例提供的一种数据处理系统的网络架构图。如图 1 所示，该网络架构可以包括业务服务器 1000 和终端设备集群，终端设备集群可以包括一个或者多个终端设备，这里将不对终端设备的数量进行限制。如图 1 所示，多个终端设备可以包括终端设备 100a、终端设备 100b、终端设备 100c、...、终端设备 100n；如图 1 所示，终端设备 100a、终端设备 100b、终端设备 100c、...、终端设备 100n 可以分别与业务服务器 1000 进行网络连接，以便于每个终端设备可以通过该网络连接与业务服务器 1000 之间进行数据交互。另外，终端设备集群 100 中的任一终端设备可以是指运行有操作系统的智能设备，本申请实施例对终端设备的操作系统不进行具体限定。

如图 1 所示的数据处理系统中的终端设备可以为智能手机、平板电脑、笔记本电脑、掌上电脑、台式计算机、移动互联网设备（MID, mobile internet device）、POS（Point Of Sales, 销售点）机、智能音箱、智能电视、智能手表、智能车载终端、虚拟现实（Virtual Reality, VR）设备、增强现实（Augmented Reality, AR）设备等，但并不局限于此。终端设备往往配置有显示装置，显示装置可以为显示器、显示屏、触摸屏等等，触摸屏可以为触控屏、触控面板等等。

如图 1 所示的数据处理系统中的业务服务器可以是单个的物理服务器，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、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。终端设备与业务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。

在一种可能的实现方式中，终端设备（如终端设备 100a）中运行有应用，如视频应用、浏览器应用、游戏应用、教育应用等，这里将不对应用进行一一举例说明。在本申请各实施例中，以视频应用为例进行

说明。对象可在终端设备中运行视频应用，而视频应用可以以 feeds 流（feed 可以是指消息来源，feed 又可译为源料、馈送、资讯提供、供稿、摘要、源、新闻订阅。其是一种资料格式，应用透过它将最新资讯传播给对象，通常以时间轴方式排列）的形式向对象提供媒体数据（如文本数据、视频数据、音频数据、图像数据等等），对象可在视频应用中浏览到这些媒体数据。对于视频应用，还可以向对象提供检索功能，对象可以通过输入某个信息的方式，来检索与该输入信息相关联的媒体数据（如文本数据、图像数据、音频数据、视频数据等）。例如，对象输入某个文本数据后，视频应用可以输出与该文本数据相关联的另一个文本数据，也可以输出与该文本数据相关联的图像数据、视频数据、音频数据等等；或者，对象输入某个图像数据后，视频应用可以输出与该图像数据相关联的另一个图像数据，也可以输出与该图像数据相关联的视频数据、文本数据、音频数据等等。而对于图像数据、文本数据、音频数据以及视频数据，可以视作不同模态类型的媒体数据，如，对于图像数据，可以理解为是模态类型为图像模态类型的媒体数据，对于视频数据，可以理解为是模态类型为视频模态类型的媒体数据。也就是说，视频应用可以提供多模态类型的媒体数据的检索功能，而为了提升多模态的检索结果的准确性，本申请提供一种基于模态路由分发机制的多模态联合训练框架，在只需训练一个模态表征模型的基础上，同时完成跨模态（跨模态也就是指两个或两个以上模态的模态组合）与单模态的模态表征特征确定的任务，基于各种模态的模态表征特征（可包含跨模态的模态表征特征与单模态的模态表征特征），即可进行不同模态类型的媒体数据的检索，由此可以减少模型存储大、检索耗时以及重复计算等问题，使得单模态与跨模态任务能够互相促进，分别保持良好地性能，提升通用性与准确性。

具体的，可以在业务服务器 1000 中部署模态表征模型，而业务服务器 1000 可以采集到对象在视频应用中所输入的信息（如输入的文本数据），基于对象在视频应用中所输入的信息，业务服务器 1000 可以通过模态表征模型对对象所输入的信息进行推理分析，以得到输入信息的模态表征向量，基于该输入信息的模态表征向量，业务服务器 1000 即可确定出与该输入信息相关联的媒体数据（如文本数据、图像数据、视频数据、音频数据等等）。其中，可以理解的是，部署于业务服务器 1000 中的模态表征模型，可以包含有不同的模态路由层，例如，具体可以包含用于对文本数据进行路由处理的文本模态路由层、用于对图像数据进行路由处理的图像模态路由层、用于对音频数据进行路由处理的音频模态路由层，还可以包含用于对多模态媒体数据进行路由处理的跨模态路由层。那么对于不同模态类型的输入信息，可以由相应的模态路由层来进行相应的路由处理（这里的路由层可以是指全连接层，路由处理可以是指通过全连接层进行线性变换处理，而由于通过全连接层进行线性变换处理实际上是指通过全连接层进行特征整合预测处理，那么这里的路由处理也可以理解为是进行特征预测处理），以得到输入信息的模态表征向量。例如，输入信息为文本数据时，可以由文本模态路由层对其进行路由处理，输入信息为图像数据时，可以由图像模态路由层对其进行路由处理。

也就是说，本申请在模态表征模型中，设置了路由分发机制，为不同模态类型设置了不同的模态路由层，对于不同模态类型的输入信息，可以分发到相应的模态路由层来进行路由处理，相较于传统技术中由统一的路由层来对不同模态类型的输入数据进行统一的处理，设置路由分发机制的方式可以提升路由层的针对性，一种模态类型的输入数据可以由对应的模态路由层来进行针对性地路由处理，基于不同的模态路由层可以适应于各种模态类型的输入数据，这样可以很好地提升模态表征模型对不同模态类型的通用性，而且，由于不同模态类型的输入数据是由相应的模态路由层进行针对性处理，那么其处理过程是符合模态类型自身特性的，所得到的处理结果（即输入数据的表征结果）也可以更为准确，即：可以提升不同模态类型的数据的表征结果的准确性。

本申请中，为了提升模态表征模型输出结果的准确性，可以对模态表征模型进行训练优化，而在模态表征模型的训练过程中，由于设置了不同的模态路由层，对于不同的模态路由层，可以采用不同模态类型的训练样本来进行针对性地训练。例如，对于文本模态路由层，可以采用文本样本数据来进行训练，对于

图像模态路由层,可以采用图像样本数据来进行训练,对于跨模态路由层,可以采用跨模态的样本数据(如既包含图像样本数据又包含文本样本数据的样本数据)来进行训练。对于文本模态路由层、图像模态路由层这种仅针对一种模态类型的路由层,本申请可以称之为单模态路由层,而对于针对多种模态类型的路由层,本申请可称之为跨模态路由层,跨模态路由层可以对多模态类型的数据进行处理。而对于单模态路由层,可以由所属模态类型为单模态类型(如文本模态类型、图像模态类型、音频模态类型等)的训练数据(训练任务组)来对其进行训练,对于跨模态路由层,可以由所属模态类型为跨模态类型的训练数据(训练任务组)来对其进行训练。

具体的,本申请可以配置训练任务组集合,例如,可以将一个样本媒体数据与其对应的模态类型,组成一个训练任务组,也可以将两个不同模态类型的样本媒体数据,与其对应的模态类型,组成一个训练任务组,基于此,通过样本媒体数据与模态类型的组合,即可得到包含单模态类型与跨模态类型的训练任务组集合。例如,样本媒体数据为样本文本数据1时,其模态类型为文本模态类型,那么将其进行组合后可以得到一个训练任务组为<样本文本数据1,文本模态类型>,该训练任务组仅包含一个文本数据,该训练任务组所属模态类型为文本模态类型(单模态类型);又如,样本媒体数据为样本图像数据1时,其模态类型为图像模态类型,那么将其进行组合后可以得到一个训练任务组为<样本图像数据1,图像模态类型>,该训练任务组仅包含一个文本数据,该训练任务组所属模态类型为图像模态类型(单模态类型);又如,样本媒体数据为样本文本数据1与样本图像数据1时,其模态类型包含文本模态类型与图像模态类型,那么将其进行组合后可以得到一个训练任务组为<样本文本数据1-样本图像数据1,文本模态类型-图像模态类型>,该训练任务组同时包含一个文本数据与一个图像数据,该训练任务组所属模态类型为跨模态类型。也就是说,本申请中的训练任务组可以包含有单模态类型的训练任务组与跨模态类型的训练任务组,对于单模态类型的训练任务组而言,其仅会包含一种模态类型的样本媒体数据,此模态类型可以为文本模态类型、图像模态类型、音频模态类型等;对于跨模态类型的训练任务组而言,其需要包含有两种或两种以上的不同模态类型的样本媒体数据。

进一步地,对于每一个训练任务组,可以调用模态表征模型对其进行注意力交互处理,其中,这里的注意力交互处理可以是指基于多头自注意力机制,对其进行多头自注意力处理,由于经过多头自注意力处理可以使得输入数据中的每一个元素与其他元素进行充分交互联系,所以本申请将多头自注意力处理称为注意力交互处理,通过注意力交互处理所得到的结果称为训练任务组所对应的注意力表征向量。随后,对于每一个注意力表征向量,可以基于训练任务组所属的模态类型,在模态表征模型所包含的单模态路由层与跨模态路由层中,确定出与其相匹配的目标路由层,通过该目标路由层,即可对训练任务组的注意力表征向量进行路由处理(特征预测处理),最终即可得到训练任务组对应的模态表征向量(训练任务组对应的模态表征向量也就是指目标路由层所输出的用于表征模态类型的特征向量,由于训练阶段需要对模态表征模型进行训练,那么目标路由层所输出的模态表征向量可以理解为一个预测值,所输出的模态表征向量即可称为预测模态表征向量,需要基于该预测模态表征向量对目标路由层进行训练优化,由此使得模态表征模型所输出的模态表征向量越来越准确)。举例而言,假设某个训练任务组所属模态类型为单模态类型中的文本模态类型,那么在得到该训练任务组的注意力表征向量后,可以在模态表征模型所包含的单模态路由层中,将文本模态路由层确定为该训练任务组的目标路由层,基于该文本模态路由层即可输出该训练任务组对应的预测模态表征向量,而基于该训练任务组对应的预测模态表征向量,可以相应地对文本模态路由层进行训练优化。

也就是说,当确定出训练任务组集合中,每个训练任务组分别对应的预测模态表征向量时,基于所属模态类型为单模态类型(如文本模态类型、图像模态类型、音频模态类型或视频模态类型)的训练任务组所对应的预测模态表征向量,可以对对应的单模态路由层(如文本模态路由层、图像模态路由层、音频模态路由层或视频模态路由层)进行优化,而优化后的单模态路由层,可以对单模态类型的数据进行路由处

理；同理，基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，可以对跨模态路由层进行优化，优化后的跨模态路由层，可以对跨模态类型的数据进行路由处理。

应当理解，本申请实施例提供一种基于模态路由分发机制的多模态联合训练框架，在只需训练一个模态表征模型的基础上，可以同时完成跨模态与单模态的模态表征向量输出的任务，基于各个模态的模态表征特征，即可进行不同模态类型的媒体数据的检索，由此可以减少模型存储大、检索耗时以及重复计算等问题，使得单模态与跨模态任务能够互相促进，分别保持良好地性能，提升通用性与准确性。

可以理解的是，本申请实施例所提供的方法可以由计算机设备执行，计算机设备包括但不限于图 1 中所提及的终端设备或业务服务器。

需要说明的是，在本申请的具体实施方式中，涉及到用户信息、用户数据（如上述对象所输入的信息）等相关的数据，均是需要经过用户手动授权许可（即经过用户同意）才进行获取到的。也就是说，当本申请以上实施例运用到具体产品或技术中时，本申请实施例所提供的方法与相关功能是在获得用户许可或者同意下所运行的（可以由用户主动开启本申请实施例所提供的功能），且相关数据的收集、使用和处理需要遵守相关地区的相关法律法规和标准。

为便于理解，接下来将结合附图对本申请实施例提供的数据处理方法进行详细描述。请参见图 2，图 2 是本申请实施例提供的一种数据处理方法的流程示意图。其中，该方法可以由计算机设备执行，该计算机设备可以为终端设备（例如，上述图 1 所示的终端设备集群中的任一终端设备，如终端设备 100a），该计算机设备也可以为服务器（如上述图 1 所对应实施例中的业务服务器 1000），该计算机设备还可以由终端设备和服务器。为便于理解，本实施例以该方法由终端设备执行为例进行说明。如图 2 所示，该数据处理方法至少可以包括以下步骤 S101-步骤 S104：

步骤 S101，获取训练任务组集合；训练任务组集合中包括训练任务组 S_i ， i 为正整数；训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在训练任务组 S_i 所属模态类型为单模态类型时，训练任务组 S_i 包含一个样本媒体数据；在训练任务组 S_i 所属模态类型为跨模态类型时，训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据。

本申请中，单模态类型可以是指媒体数据所属的模态类型，具体可以包含文本模态类型、图像模态类型、音频模态类型以及视频模态类型，某个训练任务组所属的模态类型可以基于训练任务组所包含的样本媒体数据所确定，例如，训练任务组仅包含一个样本媒体数据，且该样本媒体数据所属的模态类型为文本模态类型（即该样本媒体数据为样本文本数据），那么该训练任务组所属的模态类型可为单模态类型，且具体可为文本模态类型；又如，训练任务组仅包含一个样本媒体数据，且该样本媒体数据所属的模态类型为图像模态类型（即该样本媒体数据为样本图像数据），那么该训练任务组所属的模态可为单模态类型，且具体为图像模态类型。也就是说，当某个训练任务组仅包含一种模态类型的样本媒体数据时，该训练任务组所属的模态类型即可为单模态类型，且具体的模态类型是基于所包含的样本媒体数据的模态类型所确定的。同理可推导知，在某个训练任务组包含两个或两个以上的样本媒体数据，且这些样本媒体数据所属的模态类型不同时，即可确定该训练任务组所属的模态类型为跨模态类型（或多模态类型）。例如，某个训练任务组既包含样本文本数据（模态类型为文本模态类型的样本媒体数据），又包含有样本视频数据（模态类型为视频模态类型的样本媒体数据），那么即可确定该训练任务组所属的模态类型为跨模态类型。

也就是说，本申请中的样本媒体数据，可以是指作为训练样本的媒体数据，其具体可以包含样本文本数据、样本图像数据、样本音频数据或样本视频数据，当某个训练任务组由一种模态类型的样本媒体数据所组成时，可以确定该训练任务组所属的模态类型为单模态类型，且可以将该样本媒体数据所属的模态类型确定为该训练任务组所属的模态类型；而当某个训练任务组由多种（两种及两种以上）模态类型的样本媒体数据所组成时，可以确定该训练任务组所属的模态类型为跨模态类型或多模态类型。

具体的，对于获取训练任务组集合的具体实现方式可为：可以获取 N （ N 为正整数）个样本媒体数据；

其中，N个媒体数据中的一个样本媒体数据所属的模态类型，为第一模态类型与第二模态类型中的任意一种；且第一模态类型与第二模态类型不同；基于N个样本媒体数据中，每个样本媒体数据分别所属的模态类型，可以对N个样本媒体数据进行任务组构建处理，由此可以得到训练任务组集合。

可以理解的是，这里的第一模态类型与第二模态类型可以是指媒体数据不同的模态类型，第一模态类型与第二模态类型均可以是指文本模态类型、图像模态类型、音频模态类型或视频模态类型，且第一模态类型与第二模态类型不同，例如，第一模态类型为文本模态类型时，第二模态类型可以是指图像模态类型、音频模态类型或视频模态类型中的任意一种或多种，第一模态类型为图像模态类型时，第二模态类型可以是指文本模态类型、音频模态类型或视频模态类型中的任意一种或多种。本申请中预先获取到不同模态类型的多个作为样本的样本媒体数据，随后再基于各个样本媒体数据所属的模态类型，来进行任务组构建处理，由此来得到包含不同样本媒体数据的训练任务组，进而得到训练任务组集合。

具体的，对于基于N个样本媒体数据中，每个样本媒体数据分别所属的模态类型，对N个样本媒体数据进行任务组构建处理，得到训练任务组集合的具体实现方式可为：可以将N个样本媒体数据的第一样本媒体数据，与用于表征第一模态类型的第一标识进行组合，由此可以得到所属模态类型为第一模态类型的第一训练任务组；其中，第一样本媒体数据是指N个样本媒体数据中，所属模态类型为第一模态类型的任一样本媒体数据；应当理解的是，N个样本媒体数据中包含有第一模态类型的媒体数据与第二模态类型的媒体数据，可以将所属模态类型为第一模态类型的任一媒体数据，均称之为第一样本媒体数据，那么对于任意一个第一样本媒体数据而言，可以将其于第一模态类型的标识进行组合，由此可以得到一个训练任务组（此训练任务组可以称之为第一训练任务组），由于第一样本媒体数据的数量可能并非为一个，所以第一训练任务组的数量也可能为多个；同理，可以将N个样本媒体数据中的第二样本媒体数据，与用于表征第二模态类型的第二标识进行组合，由此可以得到所属模态类型为第二模态类型的第二训练任务组；其中，第二样本媒体数据是指N个样本媒体数据中，所属模态类型为第二模态类型的任一样本媒体数据；应当理解的是，N个样本媒体数据中包含有第一模态类型的媒体数据与第二模态类型的媒体数据，可以将所属模态类型为第二模态类型的任一媒体数据，均称之为第二样本媒体数据，那么对于任意一个第二样本媒体数据而言，可以将其于第二模态类型的标识进行组合，由此可以得到一个训练任务组（此训练任务组可以称之为第二训练任务组），由于第二样本媒体数据的数量可能并非为一个，所以第二训练任务组的数量也可能为多个；基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道，可以对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理，得到所属模态类型为跨模态类型的训练任务组；进一步地，可以将第一训练任务组与第二训练任务组，均确定为所属模态类型为单模态类型的训练任务组，并将所属模态类型为所述跨模态类型的训练任务组，与所属模态类型为所述单模态类型的训练任务组所组成的集合，确定为训练任务组集合。

可以理解的是，以第一模态类型为文本模态类型，第二模态类型包含图像模态类型与视频模态类型为例，N个样本媒体数据可以包含样本文本数据1（模态类型为文本模态类型）、样本文本数据2（模态类型为文本模态类型）、样本图像数据1（模态类型为图像模态类型）、样本视频数据1（模态类型为视频模态类型）、样本视频数据2（模态类型为视频模态类型）。其中，用于表征文本模态类型的标识可为“text”，用于表征图像模态类型的标识可为“vision”，用于表征视频模态类型的标识可为“video”，那么可以将样本文本数据1以及用于表征文本模态类型的标识“text”进行组合，由此可以得到一个训练任务组为<样本文本数据1, text>，也可以将样本文本数据1以及用于表征文本模态类型的标识“text”进行组合，由此可以得到一个训练任务组为<样本文本数据2, text>，应当理解，由于训练任务组<样本文本数据1, text>与<样本文本数据2, text>均只包含了文本模态类型的样本媒体数据，则这两个训练任务组所属的模态类型可以为单模态类型，所属的模态类型具体可为文本模态类型；同理，可以将样本图像数据1与用于表征图像模态类型的标识“vision”进行组合，由此可以得到一个训练任务组为<样本图像数据1, vision>，该训练

任务组只包含了图像模态类型的样本媒体数据,则这个训练任务组所属的模态类型可以为单模态类型,且所属的模态类型具体可为图像模态类型;同理,可以将样本视频数据1与样本视频数据2分别与用于表征视频模态类型的标识“video”进行组合,由此可以得到训练任务组<样本视频数据1, video>与<样本视频数据2, video>,这两个训练任务组均只包含了视频模态类型的样本媒体数据,则这个训练任务组所属的模态类型可以为单模态类型,且所属的模态类型具体可为视频模态类型。也就是说,对于样本媒体数据,可以将某种模态类型的样本媒体数据,与其对应的模态类型的标识进行组合,由此即可得到模态类型为单模态类型的训练任务组,同样,也可以将不同模态类型的样本媒体数据进行跨模态组合,由此即可得到模态类型为多模态类型(或称为跨模态类型)的训练任务组。换言之,所属模态类型为单模态类型的一个训练任务组,其应当包含一个样本媒体数据与该样本媒体数据所属的模态类型的标识。

具体的,可以将媒体来源渠道(媒体来源渠道可以是指样本媒体数据的获取方式或获取位置)相同的不同模态类型的样本媒体数据进行组合,以第一样本媒体数据与第二样本媒体数据的数量均为一个或多个,一个或多个第一样本媒体数据中包括第一样本媒体数据 M_j (j 为正整数),所属模态类型为跨模态类型的训练任务组,包含第一样本媒体数据 M_j 对应的训练任务组为例,对于基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道,对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理,得到所属模态类型为跨模态类型的训练任务组的具体实现方式可为:可以将第一样本媒体数据 M_j 对应的媒体来源渠道,确定为目标媒体来源渠道;随后,可以将一个或多个第二样本媒体数据中,媒体来源渠道为目标媒体来源渠道的第二样本媒体数据,确定为第一样本媒体数据 M_j 对应的关联样本媒体数据;最后可以将第一样本媒体数据、关联样本媒体数据、第一标识以及第二标识进行组合,即可得到第一样本媒体数据 M_j 对应的训练任务组。

举例而言,假设 N 个样本媒体数据包含样本文本数据a、样本文本数据b、样本图像数据a、样本视频数据h1、样本视频数据h2,对于样本文本数据a与样本图像数据a均来自于样本视频数据h1(样本文本数据a为样本视频数据h1的视频描述信息,样本图像数据a是样本视频数据h1的某个画面帧),那么可以将样本文本数据a与样本图像数据a确定为是互相关联的(样本文本数据a为样本图像数据a的关联媒体数据,同理,样本图像数据a也是样本文本数据a的关联媒体数据),样本文本数据a与样本视频数据h1也是互相关联的,样本图像数据a与样本视频数据h1也是互相关联的,基于此,可以将样本文本数据a与样本图像数据a及其对应的标识(用于表征文本模态类型的标识“text”与用于表征图像模态类型的标识“vision”)进行组合,得到训练任务组为<样本文本数据a-样本图像数据a, text-vision>,也可以将样本文本数据a与样本视频数据h1及其对应的标识(用于表征文本模态类型的标识“text”与用于表征视频模态类型的标识“video”)进行组合,得到训练任务组为<样本文本数据a-样本视频数据h1, text-video>,也可以将样本图像数据a与样本视频数据h1及其对应的标识(用于表征图像模态类型的标识“vision”与用于表征视频模态类型的标识“video”)进行组合,得到训练任务组为<样本图像数据a-样本视频数据h1, vision-video>。可见,所属模态类型为跨模态类型的一个训练任务组,其应当包含两个及两个以上的样本媒体数据以及各个样本媒体数据所属的模态类型的标识。

综上所述,本申请中的任务组构建处理是指按照各个样本媒体数据所属的模态类型、各个样本媒体数据所属模态类型的标识,将样本媒体数据与模态类型的标识进行相应组合,以得到所属模态类型为单模态类型的训练任务组与跨模态类型的训练任务组,本申请在组合样本媒体数据得到训练任务组时,各个训练任务组的格式均是统一的,无论是单模态类型的训练任务组还是多模态类型的训练任务组,均是由样本媒体数据及其对应的用于表征模态类型的标识所组合得到的,通过用于表征模态类型的标识,即可清晰地表明该训练任务组包含的模态类型是单模态类型还是跨模态类型,进而可以反映该训练任务组所属的模态类型为单模态类型还是跨模态类型。此外,本申请中的训练任务组 S_i 可以是指训练任务组集合中的任一个训练任务组,而对于训练任务组集合中的所属模态类型为单模态类型的训练任务组,其主要是由一种模

态类型（如文本模态类型、图像模态类型、视频模态类型或音频模态类型中的任一种）的样本媒体数据所组成的，且该样本媒体数据的数量为一个，那么对于该训练任务组集合中的某个训练任务组（如训练任务组 S_i ）而言，在该训练任务组为单模态类型时，该训练任务组中可仅包含一个样本媒体数据；而对于训练任务组集合中的所属模态类型为跨模态类型的训练任务组，其需要由至少两种不同的模态类型的样本媒体数据所组成，且每种模态类型下的样本媒体数据的数量需要至少一个，那么对于该训练任务组集合中的某个训练任务组（如训练任务组 S_i ）而言，在该训练任务组为跨模态类型时，该训练任务组中至少应包含有两个不同模态类型的样本媒体数据。

步骤 S102，在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量；注意力交互处理用于将训练任务组 S_i 中的不同元素进行充分交互。

本申请中，模态表征模型可以是指基于 transformer 架构的模型，通过 transformer 架构可以对每个训练任务组进行特征提取建模处理，通过特征提取建模处理可以使得训练任务组中的各个特征级的元素进行充分交互，让各个元素更能关注到与自身关联性较强的元素，所以也可以将特征提取建模处理称为注意力交互处理。具体的，transformer 架构的整个网络结构完全是由注意力机制（Attention 机制）组成，更准确地讲，transformer 架构由且仅由 self-Attention 和前馈神经网络（Feed Forward Neural Network, FFN）组成，本申请中，基于 Transformer 的可训练的神经网络（即模态表征模型）可以通过堆叠 Transformer 的形式进行搭建，采用 Attention 机制可以解决传统的 RNN（或者 LSTM, GRU 等）的计算 context 遗忘性问题。为便于理解 transformer 架构，请一并参见图 3，图 3 是本申请实施例提供的一种 transformer 架构的架构示意图。如图 3 所示，在 transformer 架构中，包含有编码器与解码器，编码器中可以包含有多头注意力层与路由层，通过多头注意力层可以将输入特征进行多头注意力处理，通过编码器中的路由层可以对多头注意力层所处理的内容进行路由处理，其中，这里的路由层可以是指一个 FFN，具体可以是指一个多层的全连接层，通过路由层可以对多头注意力层所处理的内容进行全连接处理（特征整合处理，也可以称之为路由处理）。进一步地，编码器所编码得到的特征（即路由层所输出的内容），可以输入至解码器中，通过解码器可以对编码器所编码得到的特征进行解码处理，最终解码器所解码输出的内容，可以作为输入特征所对应的模态表征向量。

基于上述可知，模态表征模型中可以包含有多头注意力层，本申请可以将其确定为多头自注意力网络层，而这里的在模态表征模型中对训练任务组 S_i 进行注意力交互处理，即可为是在多头自注意力网络层中对训练任务组 S_i 进行注意力交互处理。其具体实现方式可为：在模态表征模型中，可以通过特征提取网络层对训练任务组 S_i 所包含的样本媒体数据进行特征提取处理，由此可以得到训练任务组 S_i 对应的媒体特征；进一步地，可以通过模态表征模型中的多头自注意力网络层，对训练任务组 S_i 对应的媒体特征进行多头自注意力处理，由此即可得到训练任务组 S_i 对应的注意力表征向量。

可以理解的是，这里的特征提取网络层可以是指 embedding 层，通过 embedding 层可以提取训练任务组中，样本媒体数据的特征，得到媒体特征。其中，这里的媒体特征可以基于样本媒体数据所属的模态类型来确定，当样本媒体数据为文本模态类型时，样本媒体数据的媒体特征可以是指每一个文本词对应的词特征；当样本媒体数据为图像模态类型时，样本媒体数据的媒体特征可以是指像素特征（一个图像可以划分为多个像素网络，媒体特征可以由每一个像素网格对应的像素特征所组成）；当样本媒体数据为音频模态类型时，样本媒体数据的媒体特征可以是指音素特征；当样本媒体数据为视频模态类型时，样本媒体数据的媒体特征可以由视频帧对应的帧特征所组成。而当样本媒体数据为跨模态类型时，训练任务组的媒体特征可以由不同模态类型的样本媒体数据的媒体特征所共同组成，即可以由词特征、像素特征、音素特征、帧特征中的两种或两种以上的特征共同组成。

其中，多头自注意力网络层可以包含有 Q （ Q 为正整数）个自注意力子网络层，对于多头自注意力网络层的输出结果，可以由这些自注意力子网络层的输出结果进行融合后所得到的。具体的，以 Q 个自注

注意力子网络层包括自注意力子网络层 V_k (k 为正整数) 为例, 对于通过模态表征模型中的多头自注意力网络层, 对训练任务组 S_i 对应的媒体特征进行多头自注意力处理, 得到训练任务组 S_i 对应的注意力表征向量的具体实现方式可为: 可以获取自注意力子网络层 V_k 所包含的注意力参数矩阵, 随后可以将自注意力子网络层 V_k 所包含的注意力参数矩阵, 与训练任务组 S_i 对应的媒体特征进行运算处理, 由此可以得到注意力参数矩阵对应的线性变换矩阵; 进一步地, 可以通过自注意力子网络层 V_k 中的全连接组件, 对线性变换矩阵进行特征整合处理, 由此可得到自注意力子网络层 V_k 对应的注意力表征子向量; 当确定出 Q 个自注意力子网络层分别对应的注意力表征子向量时, 可以将 Q 个注意力表征子向量进行融合处理, 由此可以得到训练任务组 S_i 对应的注意力表征向量。

可以理解的是, 在某个自注意力子网络层中, 所包含的注意力参数矩阵可以包括 W^Q 、 W^K 以及 W^V 矩阵, 可以将训练任务组 S_i 对应的媒体特征分别与 W^Q 、 W^K 以及 W^V 进行相乘处理, 可以得到 W^Q 矩阵对应的线性变换矩阵、 W^K 矩阵对应的线性变换矩阵以及 W^V 矩阵对应的线性变换矩阵。对于线性变换矩阵 W^Q 与线性变换矩阵 W^K , 可以将其进行相乘运算处理, 对于这两者相乘运算的运算结果, 可以将其输入至全连接层 (如 Softmax 层) 进行全连接处理 (也就是特征整合处理), 对于全连接处理得到的结果, 可以与线性变换矩阵 W^V 又一次进行相乘运算处理, 与线性变换矩阵 W^V 又一次进行相乘运算处理所得到的运算结果, 即可作为一个自注意力子网络层的输出结果 (即注意力表征子向量)。每个自注意力子网络层中均包含有不同的 W^Q 、 W^K 以及 W^V 矩阵, 采用同样的原理, 可以得到不同的注意力表征子向量, 最后, 可以将这些不同的注意力表征子向量进行融合处理, 即可得到最终的注意力表征向量。

对于基于各个自注意力子网络层的注意力表征子向量, 得到最终的注意力表征向量的具体实现方式可如公式 (1) 所示:

$$E_{attn} = MSA(Q, K, V) = \text{Concat}(Head_1, Head_2 \dots Head_{H_n}) \quad \text{公式 (1)}$$

其中, E_{attn} 可用于表征最终的注意力表征向量; $Head_{H_n}$ 可用于表征自注意力子网络层 H_n 的注意力表征子向量; $\text{Concat}()$ 可用于表征拼接处理。也就是说, 可以将各个注意力表征子向量进行融合 (拼接), 融合 (拼接) 后即可得到注意力表征向量。

步骤 S103, 在模态表征模型所包含的单模态路由层与跨模态路由层中, 确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层, 基于目标路由层对训练任务组 S_i 对应的注意力表征向量进行特征预测处理, 得到训练任务组 S_i 对应的预测模态表征向量。

本申请中, 基于上述可知, 在模态表征模型所包含的编码器与解码器中, 均会包含有多头注意力层与路由层, 路由层可以将多头注意力层所输出的注意力表征向量进行特征整合处理 (即路由处理, 本申请称之为特征预测处理), 以得到最终的模态表征向量, 该模态表征向量即可是指对训练任务组进行推理理解后所得到的特征, 而为了提升对不同模态类型的数据的理解准确性, 本申请可以将模态表征模型中的路由层设置为不同的模态路由层, 将各个模态路由层与不同的模态类型对应起来, 例如, 可以将路由层设置为文本模态路由层、图像模态路由层、音频模态路由层、视频模态路由层以及跨模态路由层, 文本模态路由层可以用于对模态类型为文本模态类型的数据进行处理, 图像模态路由层可以用于对模态类型为图像模态类型的数据进行处理, 音频模态路由层可以用于对模态类型为音频模态类型的数据进行处理, 视频模态路由层可以用于对模态类型为视频模态类型的数据进行处理, 而跨模态路由层可以对模态类型为跨模态类型的数据进行处理。对于文本模态路由层、图像模态路由层、音频模态路由层、视频模态路由层, 均可以理解为是一个单模态路由层, 那么也就是说, 在模态表征模型中, 可以同时包含有单模态路由层与跨模态路由层, 那么基于此, 对于训练任务组 S_i 而言, 在确定出训练任务组 S_i 对应的注意力表征向量后, 可以在模态表征模型所包含的单模态路由层与跨模态路由层中, 确定出与训练任务组 S_i 所属的模态类型相匹配的模态路由层作为目标路由层, 再基于该目标路由层来对训练任务组 S_i 对应的注意力表征向量进行特征预测处理, 以得到该训练任务组 S_i 对应的模态表征向量 (预测模态表征向量)。

以单模态类型包含第一模态类型与第二模态类型，单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层为例，对于在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层的具体实现方式可为：若训练任务组 S_i 所属的模态类型为跨模态类型，则可以将模态表征模型中所包含的跨模态路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；若训练任务组 S_i 所属的模态类型为第一模态类型，则可以将第一模态类型对应的第一模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；而若训练任务组 S_i 所属的模态类型为第二模态类型，则可以将第二模态类型对应的第二模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层。

可以理解的是，第一模态类型与第二模态类型可以是指不同的模态类型，例如，第一模态类型可以是指文本模态类型、图像模态类型、音频模态类型以及视频模态类型中的一种或多种，而第二模态类型则可以是指文本模态类型、图像模态类型、音频模态类型以及视频模态类型中，除第一模态类型以外的剩余的模态类型，那么在训练任务组 S_i 所属的模态类型为文本模态类型时，即可将文本模态路由层确定为该训练任务组 S_i 对应的目标路由层；在训练任务组 S_i 所属的模态类型为图像模态类型时，即可将图像模态路由层确定为该训练任务组 S_i 对应的目标路由层；在训练任务组 S_i 所属的模态类型为音频模态类型时，即可将音频模态路由层确定为该训练任务组 S_i 对应的目标路由层；在训练任务组 S_i 所属的模态类型为视频模态类型时，即可将视频模态路由层确定为该训练任务组 S_i 对应的目标路由层；在训练任务组 S_i 所属的模态类型为跨模态类型时，即可将跨模态路由层确定为该训练任务组 S_i 对应的目标路由层。

为便于理解路由分发机制，请一并参见图 4，图 4 是本申请实施例提供的一种模态路由分发的架构示意图。如图 4 所示，单模态路由层具体可以为文本模态路由层、图像模态路由层、音频模态路由层或视频模态路由层，可以用于对对应的模态类型的数据进行路由处理。当输入特征为文本模态类型对应的媒体特征时，在通过多头注意力层输出注意力表征向量后，可以分发至文本模态路由层，由文本模态路由层对其进行路由处理；当输入特征为图像模态类型对应的媒体特征时，在通过多头注意力层输出注意力表征向量后，可以分发至图像模态路由层，由图像模态路由层对其进行路由处理；输入特征为音频模态类型对应的媒体特征时，在通过多头注意力层输出注意力表征向量后，可以分发至音频模态路由层，由音频模态路由层对其进行路由处理；输入特征为视频模态类型对应的媒体特征时，在通过多头注意力层输出注意力表征向量后，可以分发至视频模态路由层，由视频模态路由层对其进行路由处理；输入特征为多模态类型对应的媒体特征时，在通过多头注意力层输出注意力表征向量后，可以分发至跨模态路由层，由跨模态路由层对其进行路由处理。其中，每个模态路由层，均可以是指一个多层结构的全连接层。

步骤 S104，当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化，基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为跨模态类型的任务组进行特征预测处理。

本申请中，当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，也可以由相应模态类型的训练任务组所对应的预测模态表征向量，来对相应的模态路由层进行训练优化处理。例如，可以基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，来对单模态路由层进行优化，可以基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，来对跨模态路由层进行优化。而优化后的单模态路由层，可以用于对所属模态类型为单模态类型的任务组进行特征预测处理；优化后的跨模态路由层，可以用于对所属模态类型为跨模态类型的任务组进行特征预测处理。具体的，基于所属模态类型为文本模态类型的训练任务组所对应的预测模态表征向量，可以对文本模态路由层进行训练优化，基于所属模态类型为图像模态类型的训练任务组所对应的预测模态表征向量，可以对图像模态路

由层进行训练优化,基于所属模态类型为音频模态类型的训练任务组所对应的预测模态表征向量,可以对音频模态路由层进行训练优化,基于所属模态类型为视频模态类型的训练任务组所对应的预测模态表征向量,可以对视频模态路由层进行训练优化。训练优化后的文本模态路由层,可以对文本数据进行特征预测处理,训练优化后的图像模态路由层,可以对图像数据进行特征预测处理,训练优化后的音频模态路由层,可以对音频数据进行特征预测处理,训练优化后的视频模态路由层,可以对视频数据进行特征预测处理。

以跨模态类型的目标任务组为例,可以基于优化后的跨模态路由层对其进行处理。具体的,可以获取用于描述待分类媒体数据的目标任务组;其中,目标任务组包括至少两个不同模态类型的媒体数据;随后,可以在模态表征模型中对目标任务组进行注意力交互处理,由此可以得到目标任务组对应的注意力表征向量;进一步地,可以在优化后的跨模态路由层中,对目标任务组对应的注意力表征向量进行特征预测处理,由此可以得到目标任务组对应的预测模态表征向量;随后,可以对目标任务组对应的预测模态表征向量进行类别识别处理,由此即可得到待分类媒体数据所属的媒体类别。在确定出待分类媒体数据所属的媒体类别后,即可将该待分类媒体数据按照所属的媒体类别进行归档处理。

可以理解的是,这里是以将媒体数据进行归档的场景为例,说明了本申请实施例可以应用的场景,假设待分类媒体数据为某个视频数据,那么可以将来自于该视频数据的文本数据与图像数据作为任务组,再通过该训练优化后的跨模态路由层对其进行推理理解,以得到该视频数据的准确的模态表征向量,基于该视频数据的模态表征向量,即可确定出该视频数据所属的媒体类别(如舞蹈类别、竞赛类别、体操类别等等),通过该视频数据所属的媒体类别,即可准确地将视频数据进行归档处理。

当然,对于本申请实施例所提供的方法,应用场景并不限制于此,例如,还可以应用于多模态内容的检索场景中、媒体数据推荐的场景中等等,为便于理解,请一并参见图5,图5是本申请实施例提供的一种媒体数据检索的场景示意图。如图5所示,在视频显示界面3001中,当前向对象a曝光的视频数据为视频数据300a(该视频数据300a的创作者为名称为“啦啦爱跳舞”的对象,该视频数据300a的视频文案为“三分钟教你快速学会一支舞蹈”),在该视频显示界面3001中显示有点赞控件、评论控件以及分享控件,可以用于对象将该视频数据进行点赞、评论以及分享。在该视频显示界面3001中,还包含有搜索控件30a,可以用于对象a进行搜索相关内容。如图3所示,在对象a对该搜索控件30a产生触发操作后,终端设备可以响应这一触发操作,在该视频显示界面3001中显示输入框,对象a可以在该输入框中输入期望查阅的内容。

如图3所示,对象a在该输入框中输入文本数据“红烧鲫鱼做法”,而终端设备可以获取到该对象a所输入的文本数据“红烧鲫鱼做法”,终端设备可将该文本数据“红烧鲫鱼做法”发送至业务服务器,业务服务器可以通过训练优化后的模态表征模型,对该文本数据“红烧鲫鱼做法”进行推理理解。具体的,通过模态表征模型中的多头自注意力网络层,可以输出该文本数据“红烧鲫鱼做法”对应的注意力表征向量,可以将该注意力表征向量分发至文本模态路由层,通过该文本模态路由层,即可输出该文本数据“红烧鲫鱼做法”所对应的模态表征向量。而基于该文本数据对应的模态表征向量,即可查找到与之相关的视频数据。如图3所示,假设与之相关的视频数据为视频数据300b(该视频数据300b的创作者为名称为“美美做美食”的对象,且该视频数据300b的视频文本为“鲫鱼买来这样做,香气扑鼻!"),业务服务器可以将该视频数据300b返回至终端设备,而终端设备可以在该视频显示界面3001中显示该视频数据300b。

需要说明的是,在确定出文本数据对应的模态表征向量后,业务服务器可以基于不同向量之间的相似性,来确定两个数据之间是否是关联的。例如,对于上述文本数据“红烧鲫鱼做法”所对应的模态表征向量而言,可以先获取到候选视频数据的模态表征向量,再将文本数据“红烧鲫鱼做法”所对应的模态表征向量,与各个候选视频数据的模态表征向量分别计算相似度,相似度大于相似度阈值时,即可确定候选视频数据为文本数据“红烧鲫鱼做法”的关联视频数据。

在本申请实施例中,提供一种通用的基于模态路由分发机制的模态表征模型,可以同时为所属模态类

型为单模态类型的数据以及所属模态类型为跨模态类型的数据提供相应的数据处理功能。具体的，本申请可以采用包含单模态类型与跨模态类型的训练任务组集合，对模态表征模型进行训练，其中，在模态表征模型中，包含有单模态路由层与跨模态路由层，单模态路由层可以对所属模态类型为单模态类型的数据（注意力表征向量）进行特征预测处理，跨模态路由层可以对所属模态类型为跨模态类型的数据（注意力表征向量）进行特征预测处理，而在训练优化时，单模态路由层可以基于所属模态类型为所述单模态类型的训练任务组所对应的预测模态表征向量来进行优化，跨模态路由层可以由所属模态类型为所述跨模态类型的训练任务组所对应的预测模态表征向量来进行优化。基于此，可以使得模态表征模型中的单模态路由层与跨模态路由层能够同时进行训练优化，且在训练优化时，是基于不同模态类型的数据来分别进行训练优化，无需损失任何一种模态类型的数据，可以同时保持单模态路由层与跨模态路由层的性能，由此可以提升模态表征模型对单模态类型与跨模态类型的通用性与灵活性。综上可知，本申请可以在模态表征模型中设置路由分发机制，将路由层划分为单模态路由层与跨模态路由层，不同的路由层可以对不同模态类型的数据进行处理，在训练时也由不同模态类型的数据分别进行训练优化，每个路由层只会保留当前模态类型的相关信息，由此无需舍弃任何一种模态类型的数据，也能够同时输出跨模态类型与单模态类型的表征向量，可以很好地提升各种模态类型的数据的表征结果的准确性。综上，本申请可以在模态表征业务中，提升模态表征模型的通用性，提升模态表征结果的准确性。

进一步地，请参见图 6，图 6 是本申请实施例提供的一种对单模态路由层进行优化的流程示意图。其中，该流程可以对应于上述图 2 所对应实施例中，对于基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化的流程。该流程是以单模态类型包含第一模态类型与第二模态类型，单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层为例进行说明的流程。如图 6 所示，该流程可以至少包括以下步骤 S601-步骤 S605：

步骤 S601，将所属模态类型为第一模态类型的训练任务组，确定为第一模态训练任务组，将所属模态类型为第二模态类型的训练任务组，确定为第二模态训练任务组。

具体的，为便于区别，可以将所属模态类型为第一模态类型的训练任务组，命名为第一模态训练任务组，并将所属模态类型为第二模态类型的训练任务组，命名为第二模态训练任务组。

步骤 S602，获取第一模态训练任务组对应的真实模态表征向量，以及第二模态训练任务组对应的真实模态表征向量。

具体的，可以获取到第一模态训练任务组对应的真实模态表征向量，以及第二模态训练任务组对应的真实模态表征向量。真实模态表征向量可以是指作为训练标签的真实的模态表征向量。

步骤 S603，对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值。

具体的，可以将第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理（也就是计算出预测模态表征向量与真实模态表征向量之间的误差值），由此可以得到第一模态训练任务组对应的第一损失值。以第一模态类型为文本模态类型为例，那么第一模态训练任务组则可以为文本词序列，第一模态训练任务组对应的预测模态表征向量由文本词序列中，每个文本词分别对应的预测表征特征所组成，且第一模态训练任务组对应的真实模态表征向量，由文本词序列中，每个文本词分别对应的真实表征特征所组成，那么此时对于对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值的具体实现方式可为：可以在文本词序列所包含的文本词中获取遮掩文本词；其中，遮掩文本词是指经过遮掩处理的文本词；随后，可以在第一模态训练任务组对应的预测模态表征向量中，获取遮掩文本词对应的预测表征特征，并在第一模态训练任务组对应的真实模态表征向量中，获取遮掩文本词对应的真实表征特征；可以确定遮掩文本词对应的预测表征特征，与遮掩文本词对应的真实表征特征之

间的第一特征相似度；随后，即可将第一特征相似度确定为第一模态训练任务组对应的第一损失值。

可以理解的是，对于文本词序列，本申请中，可以采用掩蔽策略将某个文本词进行掩蔽处理，例如，可以将文本词序列中的为名词属性、动词属性或位置属性的文本词，进行掩蔽处理，而在模态表征模型所输出的预测模态表征向量中，包含有该掩蔽文本词对应的预测表征特征，在真实模态表征向量中，也包含有该掩蔽文本词对应的真实表征特征，可以计算该预测表征特征与真实表征特征之间的相似度，而该预测表征特征与真实表征特征之间的相似度，即可作为该第一模态训练任务组对应的第一损失值。

同理，当第一模态类型为图像模态类型或音频模态类型时，也可以将第一模态训练任务组中的部分元素进行掩蔽，例如，当第一模态类型为图像模态类型时，第一模态训练任务组可以为图像数据，可以将该图像数据中某些像素区域进行掩蔽处理（具体可以将相互交集比例较高的像素区域进行掩蔽），在模态表征模型所输出的预测模态表征向量中，包含有该掩蔽像素区域对应的预测表征特征，在真实模态表征向量中，也包含有该掩蔽像素区域对应的真实表征特征，同样可以计算该预测表征特征与真实表征特征之间的相似度，该预测表征特征与真实表征特征之间的相似度，即可作为该第一模态训练任务组对应的第一损失值。

需要说明的是，对于第一模态训练任务组，也可以不进行掩蔽处理，可以直接将预测模态表征向量与真实模态表征向量进行相似度计算处理，所得到的整体的相似度可以作为最终的第一损失值。还需要说明的是，除了可以通过计算相似度得到第一损失值的方式以外，还可以利用特征回归或区域分类目标的方式来计算预测值与真实值之间的损失值，本申请对于基于预测值与真实值计算损失值的方式，并不进行限制。

步骤 S604，对第二模态训练任务组对应的预测模态表征向量，与第二模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第二模态训练任务组对应的第二损失值。

具体的，对于对第二模态训练任务组对应的预测模态表征向量，与第二模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第二模态训练任务组对应的第二损失值的具体方式，可以参考上述确定第一损失值的描述，二者原理相同，这里将不再进行赘述。

步骤 S605，基于第一损失值对第一模态类型对应的第一模态子路由层进行优化，基于第二损失值对第二模态类型对应的第二模态子路由层进行优化。

具体的，基于第一损失值可以对第一模态类型对应的第一模态子路由层进行优化，而基于第二损失值可以对第二模态类型对应的第二模态子路由层进行优化。换言之，对于不同模态类型的模态子路由层，可以由相应的模态类型的预测值所得到的损失值来进行优化，具体可以是基于损失值来调整各个模态子路由层的参数，直至损失值满足收敛条件。其中，这里的收敛条件可以是指迭代次数达到迭代阈值（此迭代阈值可以基于实际的业务需求进行设定），也可以是指损失值小于损失阈值，收敛条件可以基于具体业务需求进行相应设定，本申请对其不进行限制，特此说明。

在本申请实施例中，可以在模态表征模型中设置路由分发机制，将路由层划分为单模态路由层与跨模态路由层，不同的路由层可以对不同模态类型的数据进行处理，在训练时也由不同模态类型的数据分别进行训练优化，每个路由层只会保留当前模态类型的相关信息，由此无需舍弃任何一种模态类型的数据，也能够同时输出跨模态类型与单模态类型的表征向量，可以很好地提升各种模态类型的数据的表征结果的准确性。

进一步地，请参见图 7，图 7 是本申请实施例提供的一种对跨模态路由层进行优化的流程示意图。其中，该流程可以对应于上述图 2 所对应实施例中，对于基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化的流程，该流程是以训练任务组集合中，包含至少两个所属模态类型为跨模态类型的训练任务组，至少两个所属模态类型为跨模态类型的训练任务组，包含训练任务组 S_j (j 为正整数)，训练任务组 S_j 包括所属模态类型为第一模态类型的第三样本媒体数据，以及所属模态类型为第二模态类型的第四样本媒体数据为例进行说明的流程，此时，训练任务组 S_j 所对应的预测模

态表征向量中，可以包括第三样本媒体数据对应的预测表征特征，以及第四样本媒体数据对应的预测表征特征。如图 7 所示，该流程可以至少包括以下步骤 S701-步骤 S703：

步骤 S701，确定第三样本媒体数据对应的预测表征特征，与第四样本媒体数据对应的预测表征特征之间的第二特征相似度。

具体的，对于跨模态检索的任务在于对象输入某个信息（如文本数据）后，平台可以召回与其相关联的图像/视频/音频内容。因此跨模态检索的核心思想就是进行不同模态间的对齐和一致，即匹配的图片 and 文本内容或音频内容、视频内容在特征空间最接近和相似，不能匹配的图片 and 文本内容或音频内容、视频内容在特征空间不一致。那么基于此，对于包含不同模态类型的训练任务组，需要确定第三样本媒体数据对应的预测表征特征，与第四样本媒体数据对应的预测表征特征之间的第二特征相似度，该第二特征相似度应该尽可能大，由此即可使得相匹配文本、图像、音频、视频在特征空间上是一致的。

步骤 S702，将每个所属模态类型为跨模态类型的训练任务组，均确定为跨模态训练任务组，确定训练任务组 S_j 与至少两个跨模态训练任务组中，剩余跨模态训练任务组之间的第三特征相似度；剩余跨模态训练任务组是指至少两个跨模态训练任务组中，除训练任务组 S_j 以外的跨模态训练任务组。

具体的，基于上述可知，对于媒体来源渠道不同的样本媒体数据，需要将其的相似度控制得尽可能小，那么对于训练任务组 S_j 与其他各个跨模态训练任务组，需要将其产生的第三特征相似度进行逐步减小。

步骤 S703，基于第二特征相似度与第三特征相似度，对跨模态路由层进行优化。

具体的，基于上述可知，对于第二特征相似度，本申请中，需要通过不断调整跨模态路由层的参数以将其进行极大化处理，同时，对于第三特征相似度，需要通过不断调整跨模态路由层的参数以将其进行极小化处理，直至第二特征相似度为最大值，且第三特征相似度为最小值时，确定该跨模态路由层满足了收敛条件，可以不再将跨模态路由层的参数进行调整，确定此时的跨模态路由层为最优。

在本申请实施例中，可以在模态表征模型中设置路由分发机制，将路由层划分为单模态路由层与跨模态路由层，不同的路由层可以对不同模态类型的数据进行处理，在训练时也由不同模态类型的数据分别进行训练优化，每个路由层只会保留当前模态类型的相关信息，由此无需舍弃任何一种模态类型的数据，也能够同时输出跨模态类型与单模态类型的表征向量，可以很好地提升各种模态类型的数据的表征结果的准确性。

进一步地，请参见图 8，图 8 是本申请实施例提供的一种数据处理装置的结构示意图。该数据处理装置可以是运行于计算机设备中的一个计算机程序（包括程序代码），例如该数据处理装置为一个应用软件；该数据处理装置可以用于执行图 2 所示的方法。如图 8 所示，该数据处理装置 1 可以包括：集合获取模块 11、注意力处理模块 12、路由层确定模块 13、预测模块 14、第一优化模块 15 以及第二优化模块 16。

集合获取模块 11，用于获取训练任务组集合；训练任务组集合中包括训练任务组 S_i ， i 为正整数；训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在训练任务组 S_i 所属模态类型为单模态类型时，训练任务组 S_i 包含一个样本媒体数据；在训练任务组 S_i 所属模态类型为跨模态类型时，训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

注意力处理模块 12，用于在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量；注意力交互处理用于将训练任务组 S_i 中的不同元素进行充分交互；

路由层确定模块 13，用于在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

预测模块 14，用于基于目标路由层对训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到训练任务组 S_i 对应的预测模态表征向量；

第一优化模块 15，用于当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化；

第二优化模块 16, 用于基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量, 对跨模态路由层进行优化; 优化后的单模态路由层, 用于对所属模态类型为单模态类型的任务组进行特征预测处理, 优化后的跨模态路由层, 用于对所属模态类型为跨模态类型的任务组进行特征预测处理。

其中, 集合获取模块 11、注意力处理模块 12、路由层确定模块 13、预测模块 14、第一优化模块 15 以及第二优化模块 16 的具体实现方式, 可以参见上述图 2 所对应实施例中步骤 S101-步骤 S104 的描述, 这里将不再进行赘述。

在一个实施例中, 集合获取模块 11 获取训练任务组集合的具体实现方式, 包括:

获取 N 个样本媒体数据; N 为正整数, N 个媒体数据中的一个样本媒体数据所属的模态类型, 为第一模态类型与第二模态类型中的任意一种; 第一模态类型与第二模态类型不同;

基于 N 个样本媒体数据中, 每个样本媒体数据分别所属的模态类型, 对 N 个样本媒体数据进行任务组构建处理, 得到训练任务组集合。

在一个实施例中, 集合获取模块 11 基于 N 个样本媒体数据中, 每个样本媒体数据分别所属的模态类型, 对 N 个样本媒体数据进行任务组构建处理, 得到训练任务组集合的具体实现方式, 包括:

将 N 个样本媒体数据的第一样本媒体数据, 与用于表征第一模态类型的第一标识进行组合, 得到所属模态类型为第一模态类型的第一训练任务组; 第一样本媒体数据是指 N 个样本媒体数据中, 所属模态类型为第一模态类型的任一样本媒体数据;

将 N 个样本媒体数据中的第二样本媒体数据, 与用于表征第二模态类型的第二标识进行组合, 得到所属模态类型为第二模态类型的第二训练任务组; 第二样本媒体数据是指 N 个样本媒体数据中, 所属模态类型为第二模态类型的任一样本媒体数据;

基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道, 对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理, 得到所属模态类型为跨模态类型的训练任务组;

将第一训练任务组与第二训练任务组, 均确定为所属模态类型为单模态类型的训练任务组, 将所属模态类型为所述跨模态类型的训练任务组, 与所属模态类型为所述单模态类型的训练任务组所组成的集合, 确定为训练任务组集合。

在一个实施例中, 第一样本媒体数据与第二样本媒体数据的数量均为一个或多个; 一个或多个第一样本媒体数据中包含第一样本媒体数据 M_j , j 为正整数; 所属模态类型为跨模态类型的训练任务组, 包含第一样本媒体数据 M_j 对应的训练任务组;

集合获取模块 11 基于第一样本媒体数据与第二样本媒体数据分别对应的媒体来源渠道, 对第一样本媒体数据与第二样本媒体数据进行跨模态组合处理, 得到所属模态类型为跨模态类型的训练任务组的具体实现方式, 包括:

将第一样本媒体数据 M_j 对应的媒体来源渠道, 确定为目标媒体来源渠道;

将一个或多个第二样本媒体数据中, 媒体来源渠道为目标媒体来源渠道的第二样本媒体数据, 确定为第一样本媒体数据 M_j 对应的关联样本媒体数据;

将第一样本媒体数据、关联样本媒体数据、第一标识以及第二标识进行组合, 得到第一样本媒体数据 M_j 对应的训练任务组。

在一个实施例中, 注意力处理模块 12 在模态表征模型中对训练任务组 S_i 进行注意力交互处理, 得到训练任务组 S_i 对应的注意力表征向量的具体实现方式, 包括:

在模态表征模型中, 通过特征提取网络层对训练任务组 S_i 所包含的样本媒体数据进行特征提取处理, 得到训练任务组 S_i 对应的媒体特征;

通过模态表征模型中的多头自注意力网络层, 对训练任务组 S_i 对应的媒体特征进行多头自注意力处理, 得到训练任务组 S_i 对应的注意力表征向量。

在一个实施例中，模态表征模型中的多头自注意力网络层包括 Q 个自注意力子网络层； Q 为正整数； Q 个自注意力子网络层包括自注意力子网络层 V_k ， k 为正整数；

注意力处理模块 12 通过模态表征模型中的多头自注意力网络层，对训练任务组 S_i 对应的媒体特征进行多头自注意力处理，得到训练任务组 S_i 对应的注意力表征向量的具体实现方式，包括：

获取自注意力子网络层 V_k 所包含的注意力参数矩阵，将自注意力子网络层 V_k 所包含的注意力参数矩阵，与训练任务组 S_i 对应的媒体特征进行运算处理，得到注意力参数矩阵对应的线性变换矩阵；

通过自注意力子网络层 V_k 中的全连接组件，对线性变换矩阵进行特征整合处理，得到自注意力子网络层 V_k 对应的注意力表征子向量；

当确定出 Q 个自注意力子网络层分别对应的注意力表征子向量时，将 Q 个注意力表征子向量进行融合处理，得到训练任务组 S_i 对应的注意力表征向量。

在一个实施例中，单模态类型包含第一模态类型与第二模态类型；单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层；

路由层确定模块 13 在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层的具体实现方式，包括：

若训练任务组 S_i 所属的模态类型为跨模态类型，则将模态表征模型中所包含的跨模态路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若训练任务组 S_i 所属的模态类型为第一模态类型，则将第一模态类型对应的第一模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若训练任务组 S_i 所属的模态类型为第二模态类型，则将第二模态类型对应的第二模态子路由层，确定为与训练任务组 S_i 所属的模态类型相匹配的目标路由层。

在一个实施例中，单模态类型包含第一模态类型与第二模态类型；单模态路由层包括第一模态类型对应的第一模态子路由层，与第二模态类型对应的第二模态子路由层；

第一优化模块 15 基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化的具体实现方式，包括：

将所属模态类型为第一模态类型的训练任务组，确定为第一模态训练任务组，将所属模态类型为第二模态类型的训练任务组，确定为第二模态训练任务组；

获取第一模态训练任务组对应的真实模态表征向量，以及第二模态训练任务组对应的真实模态表征向量；

对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值；

对第二模态训练任务组对应的预测模态表征向量，与第二模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第二模态训练任务组对应的第二损失值；

基于第一损失值对第一模态类型对应的第一模态子路由层进行优化，基于第二损失值对第二模态类型对应的第二模态子路由层进行优化。

在一个实施例中，第一模态类型为文本模态类型，第一模态训练任务组为文本词序列；第一模态训练任务组对应的预测模态表征向量由文本词序列中，每个文本词分别对应的预测表征特征所组成；第一模态训练任务组对应的真实模态表征向量，由文本词序列中，每个文本词分别对应的真实表征特征所组成；

第一优化模块 15 对第一模态训练任务组对应的预测模态表征向量，与第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到第一模态训练任务组对应的第一损失值的具体实现方式，包括：

在文本词序列所包含的文本词中获取遮掩文本词；遮掩文本词是指经过遮掩处理的文本词；

在第一模态训练任务组对应的预测模态表征向量中，获取遮掩文本词对应的预测表征特征，在第一模

态训练任务组对应的真实模态表征向量中，获取遮掩文本词对应的真实表征特征；

确定遮掩文本词对应的预测表征特征，与遮掩文本词对应的真实表征特征之间的第一特征相似度；

将第一特征相似度确定为第一模态训练任务组对应的第一损失值。

在一个实施例中，训练任务组集合中，包含至少两个所属模态类型为跨模态类型的训练任务组；至少两个所属模态类型为跨模态类型的训练任务组，包含训练任务组 S_j ， j 为正整数；训练任务组 S_j 包括所属模态类型为第一模态类型的第一样本媒体数据，以及所属模态类型为第二模态类型的第二样本媒体数据；训练任务组 S_j 所对应的预测模态表征向量中，包括第一样本媒体数据对应的预测表征特征，以及第二样本媒体数据对应的预测表征特征；

第二优化模块 16 基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化的具体实现方式，包括：

确定第一样本媒体数据对应的预测表征特征，与第二样本媒体数据对应的预测表征特征之间的第二特征相似度；

将每个所属模态类型为跨模态类型的训练任务组，均确定为跨模态训练任务组，确定训练任务组 S_j 与至少两个跨模态训练任务组中，剩余跨模态训练任务组之间的第三特征相似度；剩余跨模态训练任务组是指至少两个跨模态训练任务组中，除训练任务组 S_j 以外的跨模态训练任务组；

基于第二特征相似度与第三特征相似度，对跨模态路由层进行优化。

在一个实施例中，数据处理装置 1 还可以包括：任务组获取模块 17、交互处理模块 18、特征处理模块 19 以及类别识别模块 20。

任务组获取模块 17，用于获取用于描述待分类媒体数据的目标任务组；目标任务组包括至少两个不同模态类型的媒体数据；

交互处理模块 18，用于在模态表征模型中对目标任务组进行注意力交互处理，得到目标任务组对应的注意力表征向量；

特征处理模块 19，用于在优化后的跨模态路由层中，对目标任务组对应的注意力表征向量进行特征预测处理，得到目标任务组对应的预测模态表征向量；

类别识别模块 20，用于对目标任务组对应的预测模态表征向量进行类别识别处理，得到待分类媒体数据所属的媒体类别；待分类媒体数据所属的媒体类别用于对待分类媒体数据进行归档处理。

其中，任务组获取模块 17、交互处理模块 18、特征处理模块 19 以及类别识别模块 20 的具体实现方式，可以参见上述图 2 所对应实施例中步骤 S104 中的描述，这里将不再进行赘述。

在本申请实施例中，可以在模态表征模型中设置路由分发机制，将路由层划分为单模态路由层与跨模态路由层，不同的路由层可以对不同模态类型的数据进行处理，在训练时也由不同模态类型的数据分别进行训练优化，每个路由层只会保留当前模态类型的相关信息，由此无需舍弃任何一种模态类型的数据，也能够同时输出跨模态类型与单模态类型的表征向量，可以很好地提升各种模态类型的数据的表征结果的准确性。

进一步地，请参见图 9，图 9 是本申请实施例提供的一种计算机设备的结构示意图。如图 9 所示，上述计算机设备 8000 可以包括：处理器 8001，网络接口 8004 和存储器 8005，此外，上述计算机设备 8000 还包括：用户接口 8003，和至少一个通信总线 8002。其中，通信总线 8002 用于实现这些组件之间的连接通信。其中，用户接口 8003 可以包括显示屏（Display）、键盘（Keyboard），可选用户接口 8003 还可以包括标准的有线接口、无线接口。网络接口 8004 可选的可以包括标准的有线接口、无线接口（如 WI-FI 接口）。存储器 8005 可以是高速 RAM 存储器，也可以是非不稳定的存储器（non-volatile memory），例如至少一个磁盘存储器。存储器 8005 可选的还可以是至少一个位于远离前述处理器 8001 的存储装置。如图 9 所示，作为一种计算机可读存储介质的存储器 8005 中可以包括操作系统、网络通信模块、用户接口模块

以及设备控制应用程序。

在图 9 所示的计算机设备 8000 中，网络接口 8004 可提供网络通讯功能；而用户接口 8003 主要用于为用户提供输入的接口；而处理器 8001 可以用于调用存储器 8005 中存储的设备控制应用程序，以实现：

获取训练任务组集合；训练任务组集合中包括训练任务组 S_i ， i 为正整数；训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在训练任务组 S_i 所属模态类型为单模态类型时，训练任务组 S_i 包含一个样本媒体数据；在训练任务组 S_i 所属模态类型为跨模态类型时，训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

在模态表征模型中对训练任务组 S_i 进行注意力交互处理，得到训练任务组 S_i 对应的注意力表征向量；注意力交互处理用于将训练任务组 S_i 中的不同元素进行充分交互；

在模态表征模型所包含的单模态路由层与跨模态路由层中，确定与训练任务组 S_i 所属的模态类型相匹配的目标路由层，基于目标路由层对训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到训练任务组 S_i 对应的预测模态表征向量；

当确定出训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为单模态类型的训练任务组所对应的预测模态表征向量，对单模态路由层进行优化，基于所属模态类型为跨模态类型的训练任务组所对应的预测模态表征向量，对跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为跨模态类型的任务组进行特征预测处理。

应当理解，本申请实施例中所描述的计算机设备 8000 可执行前文图 2 所对应实施例中对该数据处理方法的描述，也可执行前文图 8 所对应实施例中对该数据处理装置 1 的描述，在此不再赘述。另外，对采用相同方法的有益效果描述，也不再进行赘述。

此外，这里需要指出的是：本申请实施例还提供了一种计算机可读存储介质，且上述计算机可读存储介质中存储有前文提及的数据处理的计算机设备 8000 所执行的计算机程序，且上述计算机程序包括程序指令，当上述处理器执行上述程序指令时，能够执行前文图 2 所对应实施例中对上述数据处理方法的描述，因此，这里将不再进行赘述。另外，对采用相同方法的有益效果描述，也不再进行赘述。对于本申请所涉及的计算机可读存储介质实施例中未披露的技术细节，请参照本申请方法实施例的描述。

上述计算机可读存储介质可以是前述任一实施例提供的数据处理装置或者上述计算机设备的内部存储单元，例如计算机设备的硬盘或内存。该计算机可读存储介质也可以是该计算机设备的外部存储设备，例如该计算机设备上配备的插接式硬盘，智能存储卡（smart media card, SMC），安全数字（secure digital, SD）卡，闪存卡（flash card）等。进一步地，该计算机可读存储介质还可以既包括该计算机设备的内部存储单元也包括外部存储设备。该计算机可读存储介质用于存储该计算机程序以及该计算机设备所需的其他程序和数据。该计算机可读存储介质还可以用于暂时地存储已经输出或者将要输出的数据。

本申请的一个方面，提供了一种计算机程序产品，该计算机程序产品包括计算机程序，该计算机程序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程序，处理器执行该计算机程序，使得该计算机设备执行本申请实施例中一方面提供的方法。

本申请实施例的说明书和权利要求书及附图中的术语“第一”、“第二”等是用于区别不同对象，而非用于描述特定顺序。此外，术语“包括”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、装置、产品或设备没有限定于已列出的步骤或模块，而是可选地还包括没有列出的步骤或模块，或可选地还包括对于这些过程、方法、装置、产品或设备固有的其他步骤单元。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技

术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

本申请实施例提供的方法及相关装置是参照本申请实施例提供的方法流程图和/或结构示意图来描述的，具体可由计算机程序指令实现方法流程图和/或结构示意图的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。这些计算机程序指令可提供到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或结构示意图一个方框或多个方框中指定的功能的装置。这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或结构示意图一个方框或多个方框中指定的功能。这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或结构示意图一个方框或多个方框中指定的功能的步骤。

以上所揭露的仅为本申请较佳实施例而已，当然不能以此来限定本申请之权利范围，因此依本申请权利要求所作的等同变化，仍属本申请所涵盖的范围。

权利要求书

1、一种数据处理方法，其特征在于，所述方法由计算机设备所执行，所述方法包括：

获取训练任务组集合；所述训练任务组集合中包括训练任务组 S_i ， i 为正整数；所述训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在所述训练任务组 S_i 所属模态类型为所述单模态类型时，所述训练任务组 S_i 包含一个样本媒体数据；在所述训练任务组 S_i 所属模态类型为所述跨模态类型时，所述训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

在模态表征模型中对所述训练任务组 S_i 进行注意力交互处理，得到所述训练任务组 S_i 对应的注意力表征向量；所述注意力交互处理用于将所述训练任务组 S_i 中的不同元素进行充分交互；

在所述模态表征模型所包含的单模态路由层与跨模态路由层中，确定与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层，基于所述目标路由层对所述训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到所述训练任务组 S_i 对应的预测模态表征向量；

当确定出所述训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为所述单模态类型的训练任务组所对应的预测模态表征向量，对所述单模态路由层进行优化，基于所属模态类型为所述跨模态类型的训练任务组所对应的预测模态表征向量，对所述跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为所述单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为所述跨模态类型的任务组进行特征预测处理。

2、根据权利要求 1 所述的方法，其特征在于，所述获取训练任务组集合，包括：

获取 N 个样本媒体数据； N 为正整数，所述 N 个媒体数据中的一个样本媒体数据所属的模态类型，为第一模态类型与第二模态类型中的任意一种；所述第一模态类型与所述第二模态类型不同；

基于所述 N 个样本媒体数据中，每个样本媒体数据分别所属的模态类型，对所述 N 个样本媒体数据进行任务组构建处理，得到所述训练任务组集合。

3、根据权利要求 1~2 任一项所述的方法，其特征在于，所述基于所述 N 个样本媒体数据中，每个样本媒体数据分别所属的模态类型，对所述 N 个样本媒体数据进行任务组构建处理，得到所述训练任务组集合，包括：

将所述 N 个样本媒体数据的第一样本媒体数据，与用于表征所述第一模态类型的第一标识进行组合，得到所属模态类型为所述第一模态类型的第一训练任务组；所述第一样本媒体数据是指所述 N 个样本媒体数据中，所属模态类型为所述第一模态类型的任一样本媒体数据；

将所述 N 个样本媒体数据中的第二样本媒体数据，与用于表征所述第二模态类型的第二标识进行组合，得到所属模态类型为所述第二模态类型的第二训练任务组；所述第二样本媒体数据是指所述 N 个样本媒体数据中，所属模态类型为所述第二模态类型的任一样本媒体数据；

基于所述第一样本媒体数据与所述第二样本媒体数据分别对应的媒体来源渠道，对所述第一样本媒体数据与所述第二样本媒体数据进行跨模态组合处理，得到所属模态类型为所述跨模态类型的训练任务组；

将所述第一训练任务组与所述第二训练任务组，均确定为所属模态类型为所述单模态类型的训练任务组，将所述所属模态类型为所述跨模态类型的训练任务组，与所述所属模态类型为所述单模态类型的训练任务组所组成的集合，确定为所述训练任务组集合。

4、根据权利要求 3 所述的方法，其特征在于，所述第一样本媒体数据与所述第二样本媒体数据的数量均为一个或多个；一个或多个第一样本媒体数据中包括第一样本媒体数据 M_j ， j 为正整数；所述所属模态类型为所述跨模态类型的训练任务组，包含所述第一样本媒体数据 M_j 对应的训练任务组；

所述基于所述第一样本媒体数据与所述第二样本媒体数据分别对应的媒体来源渠道，对所述第一样本媒体数据与所述第二样本媒体数据进行跨模态组合处理，得到所属模态类型为所述跨模态类型的训练任务组，包括：

将所述第一样本媒体数据 M_j 对应的媒体来源渠道，确定为目标媒体来源渠道；

将一个或多个第二样本媒体数据中，媒体来源渠道为所述目标媒体来源渠道的第二样本媒体数据，确定为所述第一样本媒体数据 M_j 对应的关联样本媒体数据；

将所述第一样本媒体数据、所述关联样本媒体数据、所述第一标识以及所述第二标识进行组合，得到所述第一样本媒体数据 M_j 对应的训练任务组。

5、根据权利要求 1~4 任一项所述的方法，其特征在于，所述在模态表征模型中对所述训练任务组 S_i 进行注意力交互处理，得到所述训练任务组 S_i 对应的注意力表征向量，包括：

在所述模态表征模型中，通过特征提取网络层对所述训练任务组 S_i 所包含的样本媒体数据进行特征提取处理，得到训练任务组 S_i 对应的媒体特征；

通过所述模态表征模型中的多头自注意力网络层，对所述训练任务组 S_i 对应的媒体特征进行多头自注意力处理，得到所述训练任务组 S_i 对应的注意力表征向量。

6、根据权利要求 5 所述的方法，其特征在于，所述模态表征模型中的多头自注意力网络层包括 Q 个自注意力子网络层； Q 为正整数；所述 Q 个自注意力子网络层包括自注意力子网络层 V_k ， k 为正整数；

所述通过所述模态表征模型中的多头自注意力网络层，对所述训练任务组 S_i 对应的媒体特征进行多头自注意力处理，得到所述训练任务组 S_i 对应的注意力表征向量，包括：

获取所述自注意力子网络层 V_k 所包含的注意力参数矩阵，将所述自注意力子网络层 V_k 所包含的注意力参数矩阵，与所述训练任务组 S_i 对应的媒体特征进行运算处理，得到所述注意力参数矩阵对应的线性变换矩阵；

通过自注意力子网络层 V_k 中的全连接组件，对所述线性变换矩阵进行特征整合处理，得到所述自注意力子网络层 V_k 对应的注意力表征子向量；

当确定出 Q 个自注意力子网络层分别对应的注意力表征子向量时，将 Q 个注意力表征子向量进行融合处理，得到所述训练任务组 S_i 对应的注意力表征向量。

7、根据权利要求 1~6 任一项所述的方法，其特征在于，所述单模态类型包含第一模态类型与第二模态类型；所述单模态路由层包括所述第一模态类型对应的第一模态子路由层，与所述第二模态类型对应的第二模态子路由层；

所述在所述模态表征模型所包含的单模态路由层与跨模态路由层中，确定与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层，包括：

若所述训练任务组 S_i 所属的模态类型为所述跨模态类型，则将所述模态表征模型中所包含的所述跨模态路由层，确定为与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若所述训练任务组 S_i 所属的模态类型为所述第一模态类型，则将所述第一模态类型对应的第一模态子路由层，确定为与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层；

若所述训练任务组 S_i 所属的模态类型为所述第二模态类型，则将所述第二模态类型对应的第二模态子路由层，确定为与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层。

8、根据权利要求 1~7 任一项所述的方法，其特征在于，所述单模态类型包含第一模态类型与第二模

态类型；所述单模态路由层包括所述第一模态类型对应的第一模态子路由层，与所述第二模态类型对应的第二模态子路由层；

所述基于所属模态类型为所述单模态类型的训练任务组所对应的预测模态表征向量，对所述单模态路由层进行优化，包括：

将所属模态类型为所述第一模态类型的训练任务组，确定为第一模态训练任务组，将所属模态类型为所述第二模态类型的训练任务组，确定为第二模态训练任务组；

获取所述第一模态训练任务组对应的真实模态表征向量，以及所述第二模态训练任务组对应的真实模态表征向量；

对所述第一模态训练任务组对应的预测模态表征向量，与所述第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到所述第一模态训练任务组对应的第一损失值；

对所述第二模态训练任务组对应的预测模态表征向量，与所述第二模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到所述第二模态训练任务组对应的第二损失值；

基于所述第一损失值对所述第一模态类型对应的第一模态子路由层进行优化，基于所述第二损失值对所述第二模态类型对应的第二模态子路由层进行优化。

9、根据权利要求 8 所述的方法，其特征在于，所述第一模态类型为文本模态类型，所述第一模态训练任务组为文本词序列；所述第一模态训练任务组对应的预测模态表征向量由所述文本词序列中，每个文本词分别对应的预测表征特征所组成；所述第一模态训练任务组对应的真实模态表征向量，由所述文本词序列中，每个文本词分别对应的真实表征特征所组成；

所述对所述第一模态训练任务组对应的预测模态表征向量，与所述第一模态训练任务组对应的真实模态表征向量，进行误差计算处理，得到所述第一模态训练任务组对应的第一损失值，包括：

在所述文本词序列所包含的文本词中获取遮掩文本词；所述遮掩文本词是指经过遮掩处理的文本词；

在所述第一模态训练任务组对应的预测模态表征向量中，获取所述遮掩文本词对应的预测表征特征，在所述第一模态训练任务组对应的真实模态表征向量中，获取所述遮掩文本词对应的真实表征特征；

确定所述遮掩文本词对应的预测表征特征，与所述遮掩文本词对应的真实表征特征之间的第一特征相似度；

将所述第一特征相似度确定为所述第一模态训练任务组对应的第一损失值。

10、根据权利要求 1~9 任一项所述的方法，其特征在于，所述训练任务组集合中，包含至少两个所属模态类型为所述跨模态类型的训练任务组；至少两个所属模态类型为所述跨模态类型的训练任务组，包含训练任务组 S_j ， j 为正整数；所述训练任务组 S_j 包括所属模态类型为第一模态类型的第三样本媒体数据，以及所属模态类型为第二模态类型的第四样本媒体数据；所述训练任务组 S_j 所对应的预测模态表征向量中，包括所述第三样本媒体数据对应的预测表征特征，以及所述第四样本媒体数据对应的预测表征特征；

所述基于所属模态类型为所述跨模态类型的训练任务组所对应的预测模态表征向量，对所述跨模态路由层进行优化，包括：

确定所述第三样本媒体数据对应的预测表征特征，与所述第四样本媒体数据对应的预测表征特征之间的第二特征相似度；

将每个所属模态类型为所述跨模态类型的训练任务组，均确定为跨模态训练任务组，确定所述训练任务组 S_j 与至少两个跨模态训练任务组中，剩余跨模态训练任务组之间的第三特征相似度；所述剩余跨模态训练任务组是指所述至少两个跨模态训练任务组中，除所述训练任务组 S_j 以外的跨模态训练任务组；

基于所述第二特征相似度与所述第三特征相似度，对所述跨模态路由层进行优化。

11、根据权利要求 1~10 任一项所述的方法，其特征在于，所述方法还包括：

获取用于描述待分类媒体数据的目标任务组；所述目标任务组包括至少两个不同模态类型的媒体数据；在所述模态表征模型中对所述目标任务组进行注意力交互处理，得到所述目标任务组对应的注意力表征向量；

在所述优化后的跨模态路由层中，对所述目标任务组对应的注意力表征向量进行特征预测处理，得到所述目标任务组对应的预测模态表征向量；

对所述目标任务组对应的预测模态表征向量进行类别识别处理，得到所述待分类媒体数据所属的媒体类别；所述待分类媒体数据所属的媒体类别用于对所述待分类媒体数据进行归档处理。

12、一种数据处理装置，其特征在于，所述装置应用于计算机设备，所述装置包括：

集合获取模块，用于获取训练任务组集合；所述训练任务组集合中包括训练任务组 S_i ， i 为正整数；所述训练任务组 S_i 所属模态类型为单模态类型或跨模态类型；在所述训练任务组 S_i 所属模态类型为所述单模态类型时，所述训练任务组 S_i 包含一个样本媒体数据；在所述训练任务组 S_i 所属模态类型为所述跨模态类型时，所述训练任务组 S_i 包含至少两个不同模态类型的样本媒体数据；

注意力处理模块，用于在模态表征模型中对所述训练任务组 S_i 进行注意力交互处理，得到所述训练任务组 S_i 对应的注意力表征向量；所述注意力交互处理用于将所述训练任务组 S_i 中的不同元素进行充分交互；

路由层确定模块，用于在所述模态表征模型所包含的单模态路由层与跨模态路由层中，确定与所述训练任务组 S_i 所属的模态类型相匹配的目标路由层；

预测模块，用于基于所述目标路由层对所述训练任务组 S_i 对应的注意力表征向量进行特征预测处理，得到所述训练任务组 S_i 对应的预测模态表征向量；

第一优化模块，用于当确定出所述训练任务组集合中每个训练任务组分别对应的预测模态表征向量时，基于所属模态类型为所述单模态类型的训练任务组所对应的预测模态表征向量，对所述单模态路由层进行优化；

第二优化模块，用于基于所属模态类型为所述跨模态类型的训练任务组所对应的预测模态表征向量，对所述跨模态路由层进行优化；优化后的单模态路由层，用于对所属模态类型为所述单模态类型的任务组进行特征预测处理，优化后的跨模态路由层，用于对所属模态类型为所述跨模态类型的任务组进行特征预测处理。

13、一种计算机设备，其特征在于，包括：处理器、存储器以及网络接口；

所述处理器与所述存储器、所述网络接口相连，其中，所述网络接口用于提供网络通信功能，所述存储器用于存储计算机程序，所述处理器用于调用所述计算机程序，以使所述计算机设备执行权利要求 1-11 任一项所述的方法。

14、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中存储有计算机程序，所述计算机程序适于由处理器加载并执行权利要求 1-11 任一项所述的方法。

15、一种计算机程序产品，其特征在于，所述计算机程序产品包括计算机程序，所述计算机程序存储在计算机可读存储介质中，所述计算机程序适于由处理器读取并执行，以使得具有所述处理器的计算机设备执行权利要求 1-11 任一项所述的方法。

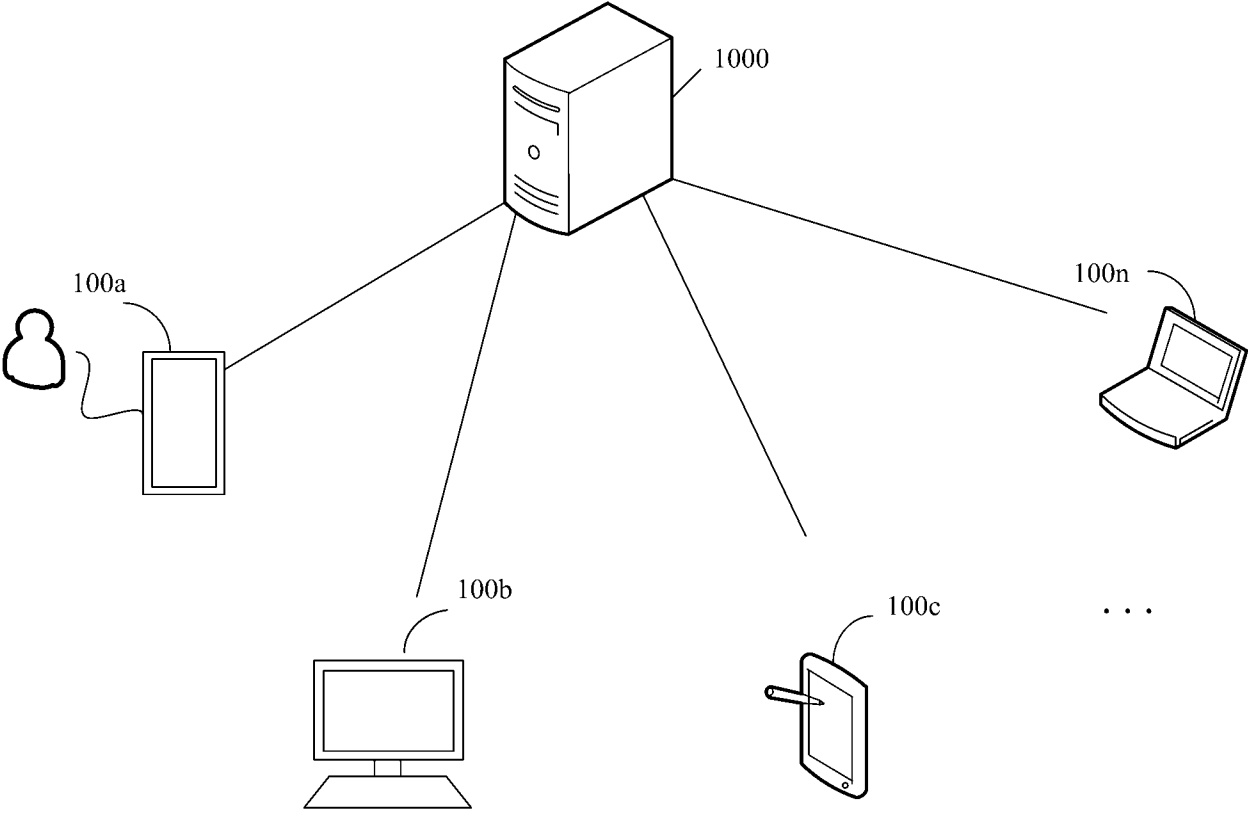


图 1

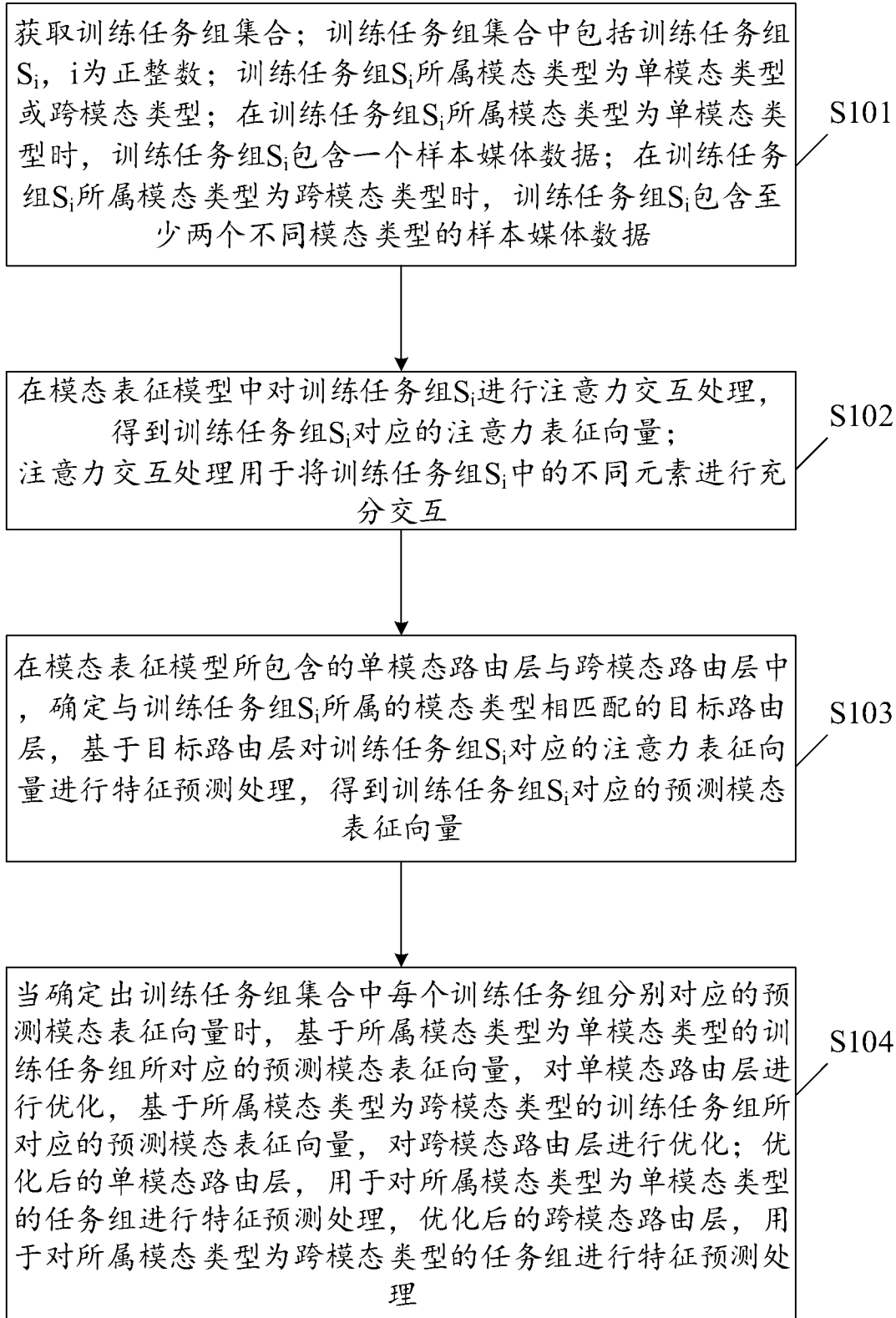


图 2

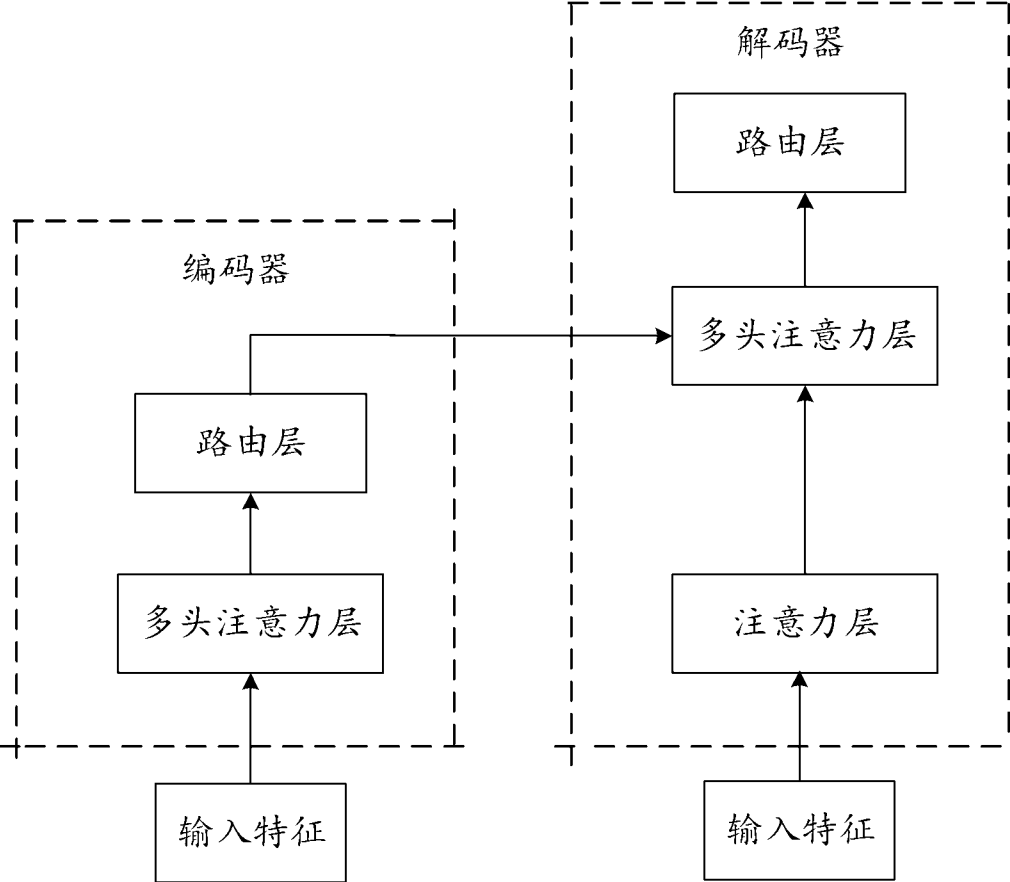


图 3

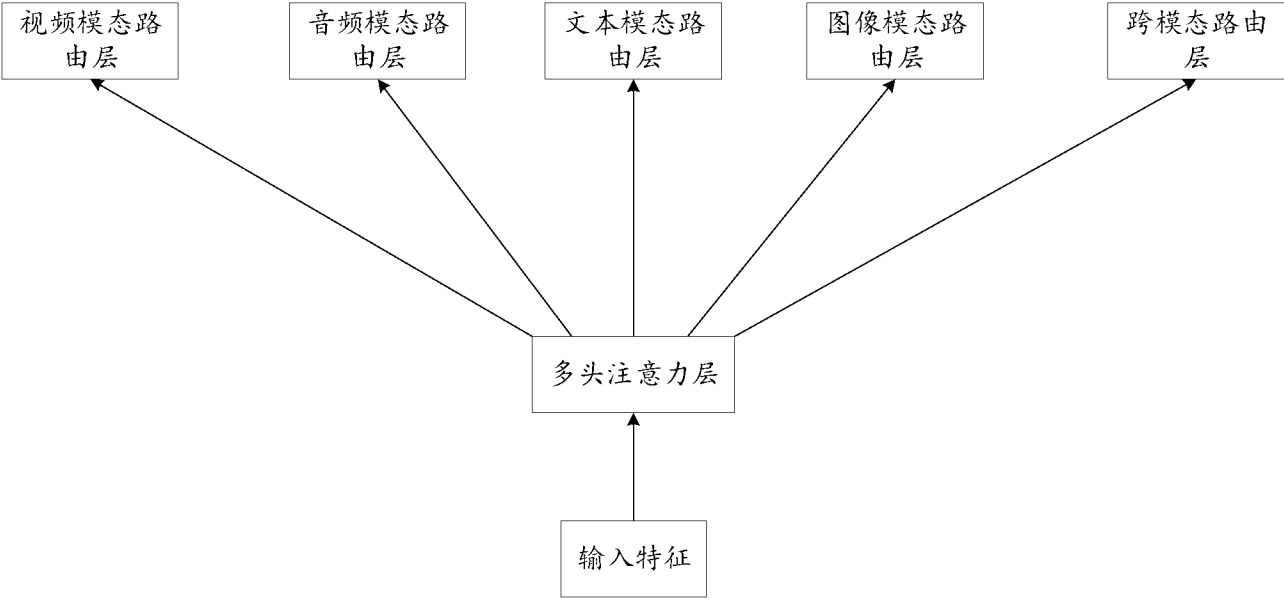


图 4

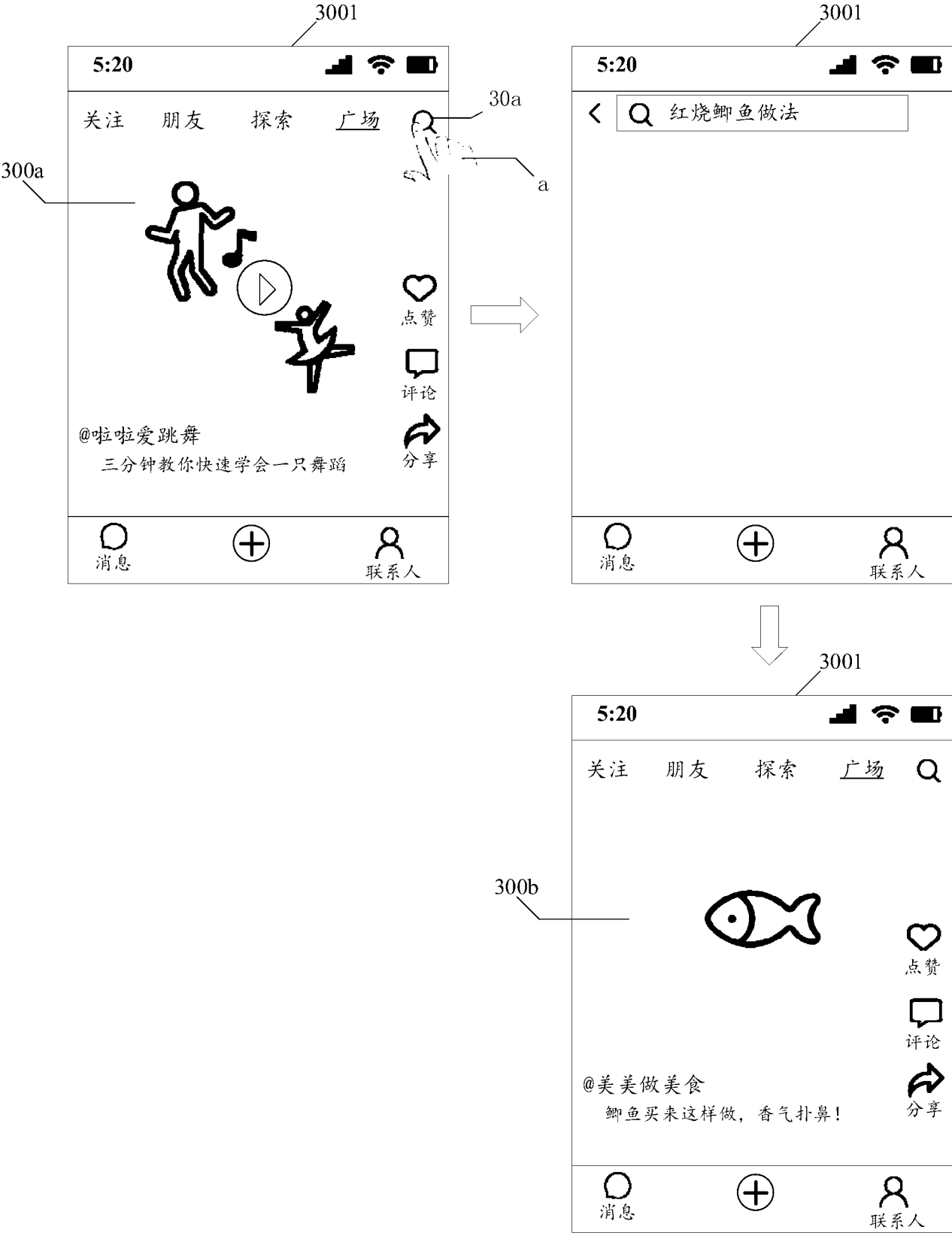


图 5

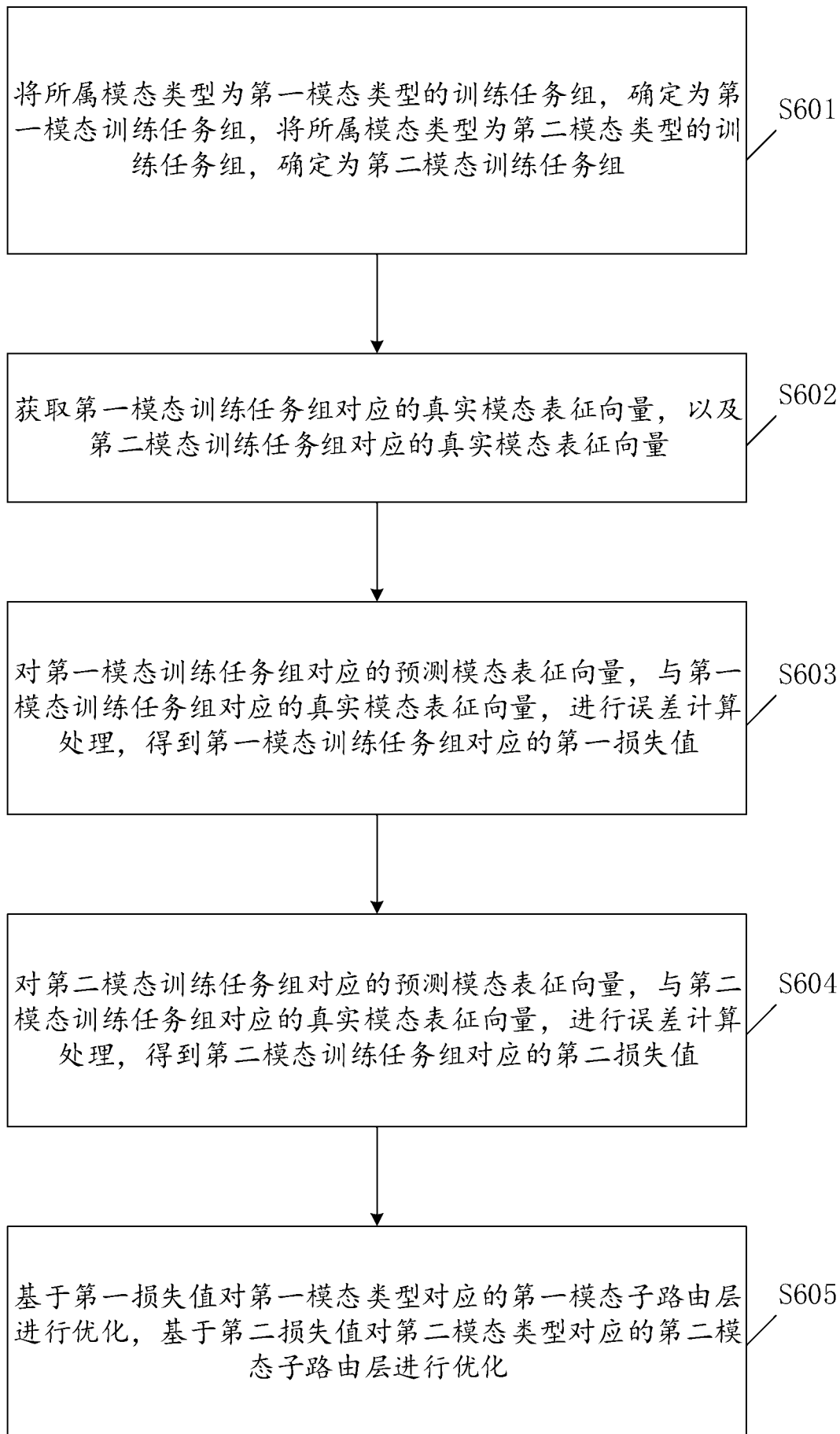


图 6

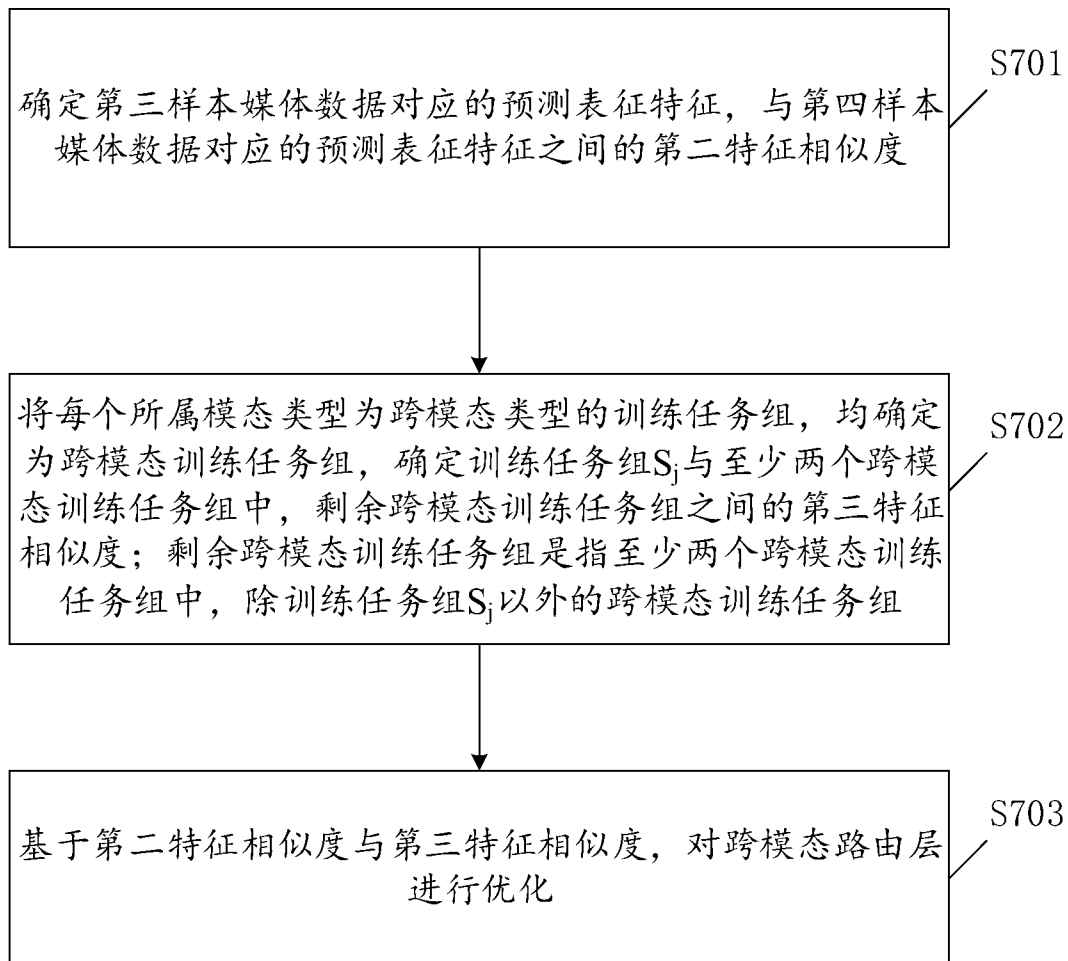


图 7

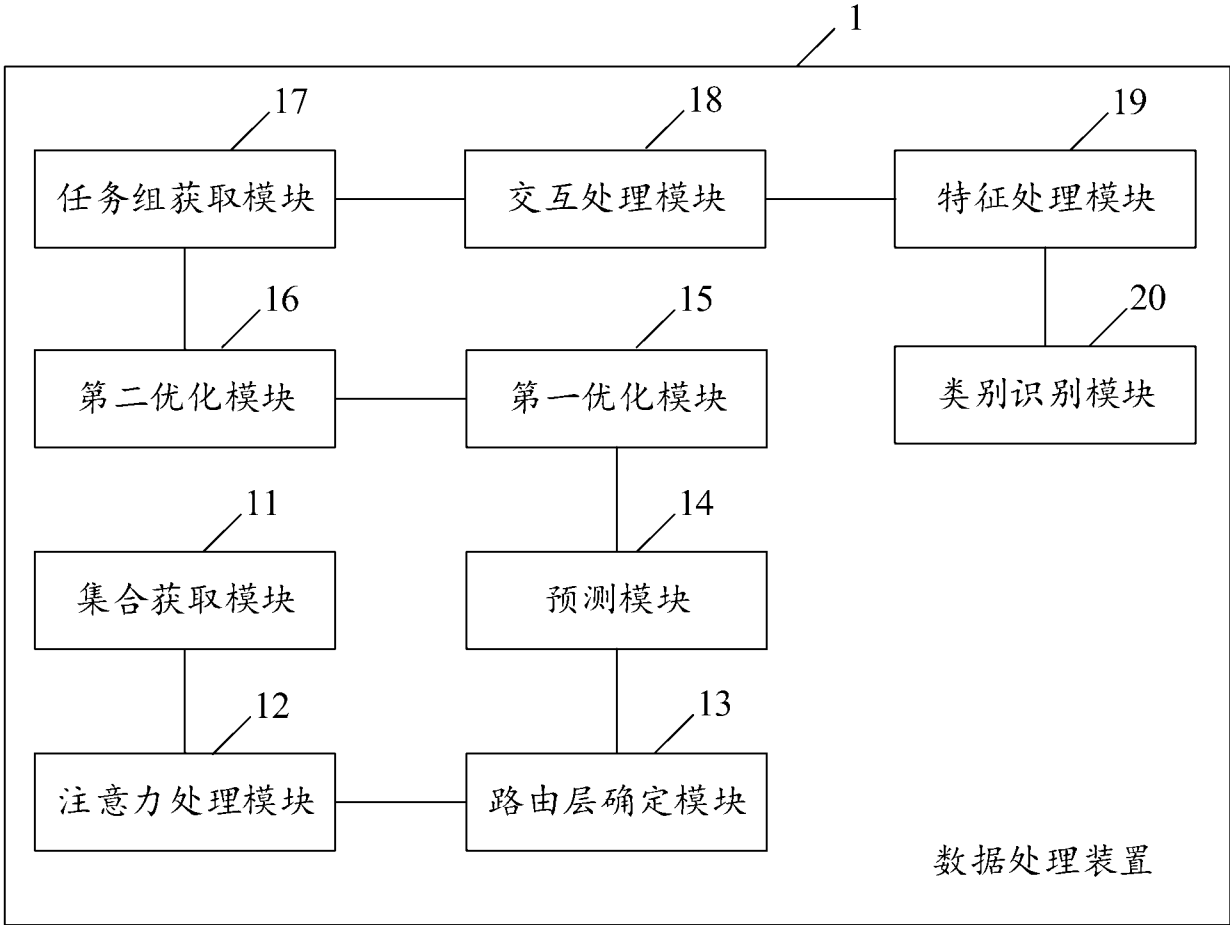


图 8

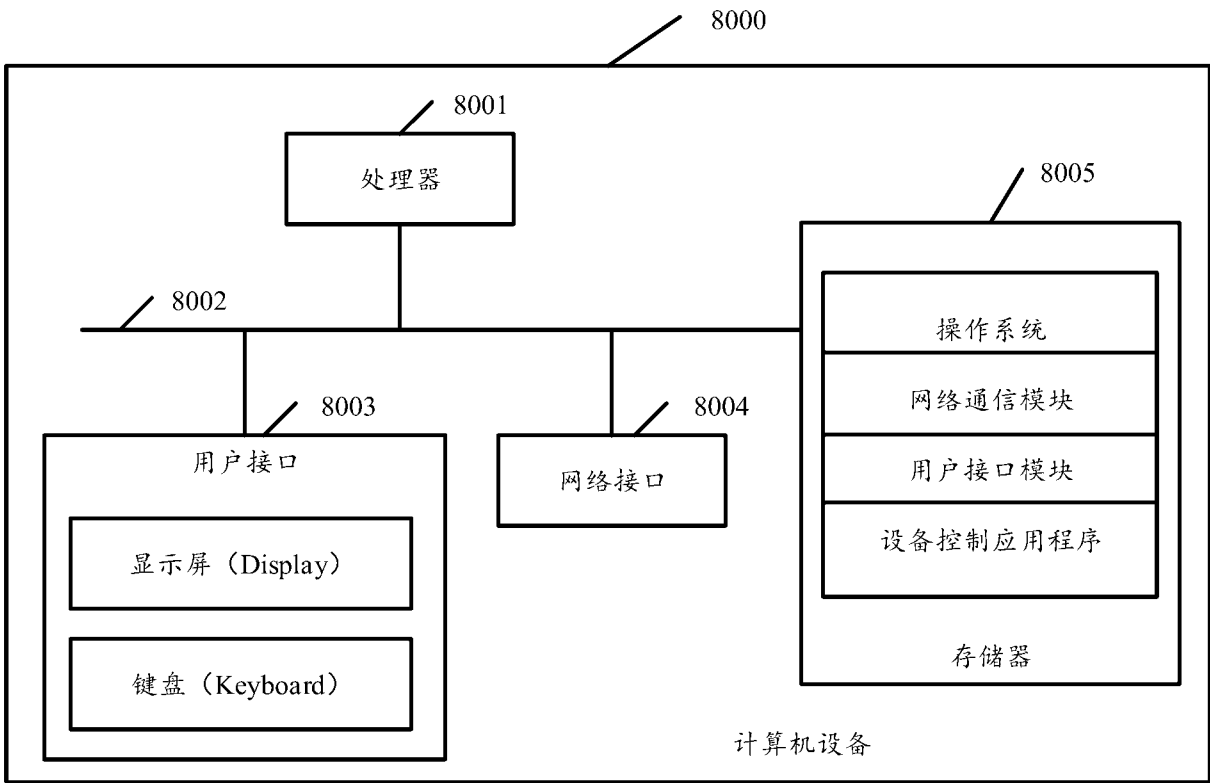


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/083332

A. CLASSIFICATION OF SUBJECT MATTER G06F16/21(2019.01)i; G06F16/23(2019.01)i; G06F16/2457(2019.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC:G06F,G06K,G06N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS, CNTXT, ENTXTC, CJFD, CNKI, DWPI, WPABS, VEN, ENTXT, IEEE: 训练, 任务, 模态, 类型, 单模, 跨模, 多模, 至少两个, 不同, 媒体, 图像, 图片, 视频, 音频, 文本, 文字, 注意力, 多头注意力, 路由层, 特征, 预测, 优化, 标签, 标识; train+, task, modality, type, single-modality, cross-modality, multi-modality, different, media, image, picture, video, audio, text, attention, transformer, routing layer, feature, predict+, optimi+, label+, identif+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116975016 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 31 October 2023 (2023-10-31) claims 1-15, description, paragraphs 6-70, and figures 1-9	1-15
X	CN 113240056 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 10 August 2021 (2021-08-10) description, paragraphs 15-52, and figures 1-7	1-15
A	CN 112668671 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 16 April 2021 (2021-04-16) entire document	1-15
A	CN 113762322 A (TENCENT TECHNOLOGY (BEIJING) CO., LTD.) 07 December 2021 (2021-12-07) entire document	1-15
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “D” document cited by the applicant in the international application “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family		
Date of the actual completion of the international search 14 May 2024		Date of mailing of the international search report 29 May 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 114419514 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 29 April 2022 (2022-04-29) entire document	1-15
A	US 2022382975 A1 (ADOBE INC.) 01 December 2022 (2022-12-01) entire document	1-15
A	US 2022392637 A1 (NEUMORA THERAPEUTICS, INC.) 08 December 2022 (2022-12-08) entire document	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2024/083332

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)	
CN	116975016	A	31 October 2023	None			
CN	113240056	A	10 August 2021	CN	113240056	B	17 May 2022
				JP	2022137145	A	21 September 2022
				US	2022327809	A1	13 October 2022
				JP	7429734	B2	08 February 2024
CN	112668671	A	16 April 2021	CN	112668671	B	24 December 2021
				US	2022292269	A1	15 September 2022
				EP	4060565	A1	21 September 2022
				KR	20220128935	A	22 September 2022
				JP	2022141587	A	29 September 2022
CN	113762322	A	07 December 2021	None			
CN	114419514	A	29 April 2022	CN	114419514	B	19 April 2024
US	2022382975	A1	01 December 2022	US	11886815	B2	30 January 2024
US	2022392637	A1	08 December 2022	WO	2022256193	A2	08 December 2022
				WO	2022256193	A3	05 January 2023

A. 主题的分类 G06F16/21(2019.01)i; G06F16/23(2019.01)i; G06F16/2457(2019.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																							
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC:G06F,G06K,G06N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS,CNXTXT,ENTXTC,CJFD,CNKI,DWPI,WPABS,VEN,ENTXT,IEEE: 训练, 任务, 模态, 类型, 单模, 跨模, 多模, 至少两个, 不同, 媒体, 图像, 图片, 视频, 音频, 文本, 文字, 注意力, 多头注意力, 路由层, 特征, 预测, 优化, 标签, 标识; train+, task, modality, type, single-modality, cross-modality, multi-modality, different, media, image, picture, video, audio, text, attention, transformer, routing layer, feature, predict+, optimi+, label+, identif+																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 116975016 A (腾讯科技(深圳)有限公司) 2023年10月31日 (2023 - 10 - 31) 权利要求1-15, 说明书第6-70段以及附图1-9</td> <td>1-15</td> </tr> <tr> <td>X</td> <td>CN 113240056 A (北京百度网讯科技有限公司) 2021年8月10日 (2021 - 08 - 10) 说明书第15-52段以及附图1-7</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 112668671 A (北京百度网讯科技有限公司) 2021年4月16日 (2021 - 04 - 16) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 113762322 A (腾讯科技(北京)有限公司) 2021年12月7日 (2021 - 12 - 07) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 114419514 A (腾讯科技(深圳)有限公司) 2022年4月29日 (2022 - 04 - 29) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>US 2022382975 A1 (ADOBE INC) 2022年12月1日 (2022 - 12 - 01) 全文</td> <td>1-15</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 116975016 A (腾讯科技(深圳)有限公司) 2023年10月31日 (2023 - 10 - 31) 权利要求1-15, 说明书第6-70段以及附图1-9	1-15	X	CN 113240056 A (北京百度网讯科技有限公司) 2021年8月10日 (2021 - 08 - 10) 说明书第15-52段以及附图1-7	1-15	A	CN 112668671 A (北京百度网讯科技有限公司) 2021年4月16日 (2021 - 04 - 16) 全文	1-15	A	CN 113762322 A (腾讯科技(北京)有限公司) 2021年12月7日 (2021 - 12 - 07) 全文	1-15	A	CN 114419514 A (腾讯科技(深圳)有限公司) 2022年4月29日 (2022 - 04 - 29) 全文	1-15	A	US 2022382975 A1 (ADOBE INC) 2022年12月1日 (2022 - 12 - 01) 全文	1-15
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 116975016 A (腾讯科技(深圳)有限公司) 2023年10月31日 (2023 - 10 - 31) 权利要求1-15, 说明书第6-70段以及附图1-9	1-15																					
X	CN 113240056 A (北京百度网讯科技有限公司) 2021年8月10日 (2021 - 08 - 10) 说明书第15-52段以及附图1-7	1-15																					
A	CN 112668671 A (北京百度网讯科技有限公司) 2021年4月16日 (2021 - 04 - 16) 全文	1-15																					
A	CN 113762322 A (腾讯科技(北京)有限公司) 2021年12月7日 (2021 - 12 - 07) 全文	1-15																					
A	CN 114419514 A (腾讯科技(深圳)有限公司) 2022年4月29日 (2022 - 04 - 29) 全文	1-15																					
A	US 2022382975 A1 (ADOBE INC) 2022年12月1日 (2022 - 12 - 01) 全文	1-15																					
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																							
国际检索实际完成的日期 2024年5月14日		国际检索报告邮寄日期 2024年5月29日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		授权官员 钱紫娟 电话号码 (+86) 027-59371331																					

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	US 2022392637 A1 (NEUMORA THERAPEUTICS INC) 2022年12月8日 (2022 - 12 - 08) 全文	1-15

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/083332

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	116975016	A	2023年10月31日	无	
CN	113240056	A	2021年8月10日	CN	113240056 B 2022年5月17日
				JP	2022137145 A 2022年9月21日
				US	2022327809 A1 2022年10月13日
				JP	7429734 B2 2024年2月8日
CN	112668671	A	2021年4月16日	CN	112668671 B 2021年12月24日
				US	2022292269 A1 2022年9月15日
				EP	4060565 A1 2022年9月21日
				KR	20220128935 A 2022年9月22日
				JP	2022141587 A 2022年9月29日
CN	113762322	A	2021年12月7日	无	
CN	114419514	A	2022年4月29日	CN	114419514 B 2024年4月19日
US	2022382975	A1	2022年12月1日	US	11886815 B2 2024年1月30日
US	2022392637	A1	2022年12月8日	WO	2022256193 A2 2022年12月8日
				WO	2022256193 A3 2023年1月5日