(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114444609 A
(43)申请公布日 2022.05.06

(21)申请号 202210118785.7

(22)申请日 2022.02.08

(71)申请人 腾讯科技(深圳)有限公司
地址 518000 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 刘世兴 王智圣 郑磊

(74)专利代理机构 北京市立方律师事务所 11330
专利代理师 张筱宁

(51)Int.Cl.
G06K 9/62(2022.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

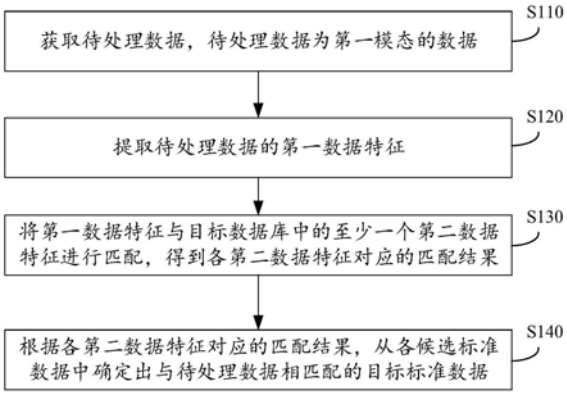
权利要求书3页 说明书24页 附图5页

(54)发明名称

数据处理方法、装置、电子设备及计算机可读存储介质

(57)摘要

本申请实施例提供了一种数据处理方法、装置、电子设备及计算机可读存储介质，涉及人工智能、多媒体、游戏及云技术领域。该方法包括：获取待处理数据，待处理数据为第一模态的数据；提取待处理数据的第一数据特征；将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，根据各第二数据特征对应的匹配结果，从各候选标准数据中确定出与待处理数据相匹配的目标标准数据，其中，目标数据库中包括至少一个候选标准数据以及每个候选标准数据的第二数据特征，候选标准数据为第二模态的数据。基于本申请实施例提供的方法，能够简单、快捷的实现不同模态的数据之间的匹配。



CN 114444609 A

1.一种数据处理方法,其特征在于,包括:

获取待处理数据,所述待处理数据为第一模态的数据;

提取所述待处理数据的第一数据特征;

将所述第一数据特征与目标数据库中的至少一个第二数据特征进行匹配,得到各所述第二数据特征对应的匹配结果,其中,所述目标数据库中包括至少一个候选标准数据以及每个所述候选标准数据的第二数据特征,所述候选标准数据为第二模态的数据;

根据各所述第二数据特征对应的匹配结果,从各所述候选标准数据中确定出与所述待处理数据相匹配的目标标准数据。

2.根据权利要求1所述的方法,其特征在于,还包括:

根据所述第一数据特征,确定所述待处理数据的数据类型;

所述将所述第一数据特征与目标数据库中的至少一个第二数据特征进行匹配,包括:

在所述待处理数据的数据类型为指定类型时,将所述第一数据特征与目标数据库中的至少一个第二数据特征进行匹配。

3.根据权利要求1所述的方法,其特征在于,所述第一模态的数据和所述第二模态的数据为不同模态的数据,所述第一模态的数据包括文本、语音、视频或图像中的至少一种,所述第二模态的数据包括文本、语音、视频或图像中的至少一种。

4.根据权利要求1至3中任一项所述的方法,其特征在于,所述候选标准数据是与标准数据库中的第一标准数据相匹配的标准表达,所述第一标准数据为第一模态的数据,一个所述第一标准数据对应至少一个标准表达。

5.根据权利要求4所述的方法,其特征在于,还包括:

在所述标准数据库存在新增第一标准数据时,获取所述新增第一标准数据对应的至少一个标准表达;

提取所述新增第一标准数据对应的各标准表达的第二数据特征;

将所述新增第一标准数据对应的各标准表达、以及各标准表达对应的第二数据特征关联存储至所述目标数据库中。

6.根据权利要求1所述的方法,其特征在于,所述第一数据特征是通过第一特征提取网络提取得到的;所述候选标准数据的第二数据特征是通过第二特征提取网络提取得到的;所述第一特征提取网络和所述第二特征提取网络是通过以下方式训练得到的:

获取训练数据集,所述训练数据集包括第一训练集,所述第一训练集中的每个第一样本包括第一模态的第一数据、以及与该第一数据相匹配的第二模态的第二数据;

基于所述训练数据集对初始的神经网络模型进行迭代训练,直至训练总损失值满足预设训练结束条件,其中,所述神经网络模型包括第一网络模型和第二网络模型,将满足所述训练结束条件时的第一网络模型作为所述第一特征提取网络,将满足所述训练结束条件时的第二网络模型作为所述第二特征提取网络;所述训练的过程包括:

将各所述第一数据输入到第一网络模型中,得到各所述第一数据的特征,将各所述第二数据输入到第二网络模型中,得到各所述第二数据的特征;

基于各所述第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度,确定第一训练损失值;其中,所述第一负例包括一个第一样本的第一数据和另一个第一样本的第二数据;

若所述第一训练损失值不满足第一预设条件，则对所述第一网络模型和第二网络模型的模型参数进行调整，所述训练总损失值满足预设训练结束条件包括所述第一训练损失值满足第一预设条件。

7.根据权利要求6所述的方法，其特征在于，所述将各所述第一数据输入到第一网络模型中，得到各所述第一数据的特征，包括：

对于每个所述第一数据，通过所述第一网络模型对该第一数据执行以下操作，得到该第一数据的特征：

将该第一数据划分为至少两个子数据，得到该第一数据对应的子数据序列；

基于词典，提取所述子数据序列中各个子数据的特征，其中，所述词典包括多个数据元素，每个所述子数据的特征包括的特征值的个数等于所述词典中元素的数量，一个特征值表征了该子数据中包含词典中与该特征值的位置相对应的数据元素的概率；

基于各所述子数据的特征，得到该第一数据的特征；

所述方法还包括：

对于每个所述第二数据，基于所述词典，确定该第二数据对应于所述词典的数据特征，该数据特征表征了该第二数据对应于词典中各个数据元素的概率；

所述确定第一训练损失值，包括：

基于各所述第一样本中的第一数据的各个子数据的特征与第二数据对应于所述词典的数据特征之间的匹配程度、各所述第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值。

8.根据权利要求6所述的方法，其特征在于，所述基于各所述第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值，包括：

确定各所述第一样本的第一数据的特征和第二数据的特征之间的差异程度，得到第一损失值；

对于每个所述第一数据，确定该第一数据对应的第一相似度以及该第一数据对应的第二相似度，其中，所述第一相似度是该第一数据的特征和与该第一数据相匹配的第二数据的特征之间的相似度，所述第二相似度是该第一数据所在的第一负例中第一数据的特征和第二数据的特征之间的相似度；

获取各所述第一数据对应的参考标签，所述参考标签包括第一相似度对应的相似度标签和第二相似度对应的相似度标签；

基于各所述第一数据所对应的预测相似度和参考标签，确定第二损失值，其中，所述预测相似度包括所述第一相似度和所述第二相似度，所述第二损失值表征了各所述第一数据所对应的预测相似度和参考标签之间的差异；

根据所述第一损失值和所述第二损失值，确定所述第一训练损失值。

9.根据权利要求6至8中任一项所述的方法，其特征在于，所述候选标准数据为指定类型的第一标准数据所对应的第二模态的标准表达；所述初始的神经网络模型还包括分类模型；

所述训练数据集还包括第二训练集，所述第二训练集中的每个第二样本包括第一模态

的第三数据、与该第三数据相匹配的第二模态的第四数据、以及该第三数据的类型标签，其中，所述第二训练集中的第三数据包括指定类型的第三数据和非指定类型的第三数据；

在得到所述第一训练损失值满足第一预设条件的神经网络模型之后，所述训练的过程还包括：

基于所述第二训练集继续对所述神经网络模型重复执行训练操作，直至第二训练损失值满足第二预设条件，其中，所述训练总损失值满足预设训练结束条件，还包括所述第二训练损失值满足第二预设条件；所述训练操作包括：

将各所述第三数据输入到第一网络模型中，得到各所述第三数据的特征，将各所述第四数据输入到第二网络模型中，得到各所述第四数据的特征，将各所述第三数据的特征的输入至分类模型中，得到各所述第三数据对应的预测类型；

基于各所述第二样本中的第三数据的特征与第四数据的特征的匹配程度、各第二负例中的第三数据的特征和第四数据的特征的匹配程度、以及各所述第三数据的类型标签和预测类型之间的匹配程度，确定第二训练损失值；

若所述第二训练损失值不满足所述第二预设条件，则对所述神经网络模型的模型参数进行调整。

10.一种数据处理装置，其特征在于，所述装置包括：

待处理数据获取模块，用于获取待处理数据，所述待处理数据为第一模态的数据；

特征获取模块，用于提取所述待处理数据的第一数据特征；

数据识别模块，用于将所述第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各所述第二数据特征对应的匹配结果，以及根据各所述第二数据特征对应的匹配结果，从各候选标准数据中确定出与所述待处理数据相匹配的目标标准数据；

其中，所述目标数据库中包括至少一个所述候选标准数据以及每个所述候选标准数据的第二数据特征，所述候选标准数据为第二模态的数据。

11.一种电子设备，其特征在于，所述电子设备包括存储器和处理器，所述存储器中存储有计算机程序，所述处理器执行所述计算机程序以实现权利要求1至9中任一项所述的方法。

12.一种计算机可读存储介质，其特征在于，所述存储介质中存储有计算机程序，所述计算机程序被处理器执行时实现权利要求1至9中任一项所述的方法。

13.一种计算机程序产品，其特征在于，所述计算机产品包括计算机程序，所述计算机程序被处理器执行时实现权利要求1至9中任一项所述的方法。

数据处理方法、装置、电子设备及计算机可读存储介质

技术领域

[0001] 本申请涉及人工智能、多媒体技术、游戏以及云技术领域，具体而言，本申请涉及一种数据处理方法、装置、电子设备及计算机可读存储介质。

背景技术

[0002] 随着语音识别技术的发展和推广，语音识别的应用已经出现在各种各样的应用场景中，比如，目前，绝大多数的电子设备都安装有人工智能AI语音助手，AI语音助手可以基于语音识别技术对采集的语音数据进行识别，得到对应的文本内容，并可以基于识别出的文本内容执行相应的功能。

[0003] 现有技术中，语音识别技术大多都是通过复杂的语音识别模型实现的，通常是通过语音编码器获取语音数据的编码特征，再通过分类网络预测语音数据的类别，虽然能够一定程度上满足需求，但是方案实现复杂、成本较高，而且方案的扩展性较差，无法新增类别，在语音数据的类别较多时，识别准确率也难以保证。

发明内容

[0004] 本申请实施例提供了一种数据处理方法、装置、电子设备及计算机可读存储介质，基于该方法能够简单、快捷的实现不同模态的数据间的匹配。本申请实施例提供的技术方案如下：

[0005] 一方面，本申请实施例提供了一种数据处理方法，该方法包括：

[0006] 获取待处理数据，待处理数据为第一模态的数据；

[0007] 提取待处理数据的第一数据特征；

[0008] 将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各第二数据特征对应的匹配结果，其中，目标数据库中包括至少一个候选标准数据以及每个候选标准数据的第二数据特征，上述候选标准数据为第二模态的数据；

[0009] 根据各第二数据特征对应的匹配结果，从各候选标准数据中确定出与待处理数据相匹配的目标标准数据。

[0010] 另一方面，本申请实施例提供了一种数据处理装置，该装置包括：

[0011] 待处理数据获取模块，用于获取待处理数据，待处理数据为第一模态的数据；

[0012] 特征获取模块，用于提取待处理数据的第一数据特征；

[0013] 数据识别模块，用于将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各第二数据特征对应的匹配结果，以及根据各第二数据特征对应的匹配结果，从各候选标准数据中确定出与待处理数据相匹配的目标标准数据；

[0014] 其中，目标数据库中包括至少一个候选标准数据以及每个候选标准数据的第二数据特征，候选标准数据为第二模态的数据。

[0015] 可选的，数据识别模块还用于：根据第一数据特征，确定待处理数据的数据类型；相应的，数据识别模块可以用于：

[0016] 在待处理数据的数据类型为指定类型时，将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配。

[0017] 可选的，第一模态的数据和第二模态的数据为不同模态的数据，第一模态的数据包括文本、语音、视频或图像中的至少一种，第二模态的数据包括文本、语音、视频或图像中的至少一种。

[0018] 可选的，候选标准数据是与标准数据库中的第一标准数据相匹配的标准表达，第一标准数据为第一模态的数据，一个第一标准数据对应至少一个标准表达。

[0019] 可选的，特征获取模块还用于：在标准数据库存在新增第一标准数据时，获取新增第一标准数据对应的至少一个标准表达；提取新增第一标准数据对应的各标准表达的第二数据特征；将新增第一标准数据对应的各标准表达、以及各标准表达对应的第二数据特征关联存储至目标数据库中。

[0020] 可选的，第一数据特征是通过第一特征提取网络提取得到的；候选标准数据的第二数据特征是通过第二特征提取网络提取得到的；其中，第一特征提取网络和第二特征提取网络是由模型训练模块通过以下方式训练得到的：

[0021] 获取训练数据集，训练数据集包括第一训练集，第一训练集中的每个第一样本包括第一模态的第一数据、以及与该第一数据相匹配的第二模态的第二数据；

[0022] 基于训练数据集对初始的神经网络模型进行迭代训练，直至训练总损失值满足预设训练结束条件，其中，神经网络模型包括第一网络模型和第二网络模型，将满足训练结束条件时的第一网络模型作为第一特征提取网络，将满足训练结束条件时的第二网络模型作为第二特征提取网络；训练的过程包括：

[0023] 将各第一数据输入到第一网络模型中，得到各第一数据的特征，将各第二数据输入到第二网络模型中，得到各第二数据的特征；

[0024] 基于各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值；其中，上述第一负例包括一个第一样本的第一数据和另一个第一样本的第二数据；

[0025] 若第一训练损失值不满足第一预设条件，则对第一网络模型和第二网络模型的模型参数进行调整，上述训练总损失值满足预设训练结束条件包括第一训练损失值满足第一预设条件。

[0026] 可选的，模型训练模块在将各第一数据输入到第一网络模型中，得到各第一数据的特征时用于：

[0027] 对于每个第一数据，通过第一网络模型对该第一数据执行以下操作，得到该第一数据的特征：

[0028] 将该第一数据划分为至少两个子数据，得到该第一数据对应的子数据序列；基于词典，提取该子数据序列中各个子数据的特征，基于各子数据的特征，得到第一数据的特征，其中，词典包括多个数据元素，每个子数据的特征包括的特征值的个数等于词典中元素的数量，一个特征值表征了该子数据中包含词典中与该特征值的位置相对应的数据元素的概率；

[0029] 模型训练模块还用于：对于每个第二数据，基于词典，确定该第二数据对应于词典的数据特征，该数据特征表征了该第二数据对应于词典中各个数据元素的概率；

[0030] 模型训练模块在确定第一训练损失值时用于:基于各第一样本中的第一数据的各个子数据的特征与第二数据对应于词典的数据特征之间的匹配程度、各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度,确定第一训练损失值。

[0031] 可选的,模型训练模块在确定第一训练损失值时可以用于:

[0032] 确定各第一样本的第一数据的特征和第二数据的特征之间的差异程度,得到第一损失值;

[0033] 对于每个第一数据,确定该第一数据对应的第一相似度以及该第一数据对应的第二相似度,其中,第一相似度是该第一数据的特征和与该第一数据相匹配的第二数据的特征之间的相似度,第二相似度是该第一数据所在的第一负例中第一数据和第二数据的相似度;

[0034] 获取各第一数据对应的参考标签,该参考标签包括第一相似度对应的相似度标签和第二相似度对应的相似度标签;

[0035] 基于各第一数据所对应的预测相似度和参考标签,确定第二损失值,其中,预测相似度包括第一相似度和第二相似度,第二损失值表征了各第一数据所对应的预测相似度和参考标签之间的差异;

[0036] 根据第一损失值和第二损失值,确定第一训练损失值。

[0037] 可选的,候选标准数据为指定类型的第一标准数据对应的第二模态的标准表达;初始的神经网络模型还包括分类模型;训练数据集还包括第二训练集,第二训练集中的每个第二样本包括第一模态的第三数据、以及与第三数据相匹配的第二模态的第四数据,其中,第二训练数据集中的第三数据包括指定类型的第三数据和非指定类型的第三数据,每个第二样本还包括该样本中的第三数据的类型标签;在得到第一训练损失值满足第一预设条件的神经网络模型之后,模型训练模块还用于执行以下训练过程:

[0038] 基于第二训练集继续对神经网络模型重复执行训练操作,直至第二训练损失值满足第二预设条件,其中,训练总损失值满足预设训练结束条件还包括第二训练损失值满足第二预设条件;上述训练操作包括:

[0039] 将各第三数据输入到第一网络模型中,得到各第三数据的特征,将各第四数据输入到第二网络模型中,得到各第四数据的特征,将各第三数据的特征的输入至分类模型中,得到各第三数据对应的预测类型;

[0040] 基于各第二样本中的第三数据的特征与第四数据的特征的匹配程度、各第二负例中的第三数据的特征和第四数据的特征的匹配程度、以及各第三数据的类型标签和预测类型之间的匹配程度,确定第二训练损失值;

[0041] 若第二训练损失值不满足第二预设条件,则对神经网络模型的模型参数进行调整。

[0042] 可选的,第一模态的数据为语音,第二模态的数据为文本,数据元素为音素。

[0043] 可选的,指定类型为指令型语音。

[0044] 另一方面,本申请实施例还提供了一种电子设备,该电子设备包括存储器和处理器,存储器中存储有计算机程序,处理器执行该计算机程序以实现本申请任一可选实施例中提供的方法。

[0045] 另一方面，本申请实施例还提供了一种计算机可读存储介质，该存储介质中存储有计算机程序，该计算机程序被处理器执行时实现本申请任一可选实施例中提供的方法。

[0046] 另一方面，本申请实施例还提供了一种计算机程序产品，该计算机产品包括计算机程序，该计算机程序被处理器执行时实现本申请任一可选实施例中提供的方法。

[0047] 本申请实施例提供的技术方案带来的有益效果如下：

[0048] 本申请实施例提供的数据处理方法，提供了一种新颖的数据处理思路，采用该方法在对待处理数据进行处理时，在获取到待处理数据的数据特征之后，可以无需对该数据特征进行复杂、繁琐的特征识别，而是可以通过特征匹配的方式，简单、快捷的实现了不同模态数据之间的匹配，可以大大减少计算量，提高数据处理效率。此外，由于目标数据库存储的都是候选标准数据，通过本申请实施例的该方法所确定出的也就是与第一模态的待处理数据相匹配的第二模态的标准数据，识别准确性亦可以得到很好的保证。

附图说明

[0049] 为了更清楚地说明本申请实施例中的技术方案，下面将对本申请实施例描述中所需要使用的附图作简单地介绍。

[0050] 图1为本申请实施例提供的一种数据处理方法的流程示意图；

[0051] 图2为本申请实施例提供的一种数据处理系统的结构示意图；

[0052] 图3为本申请实施例提供的一种神经网络模型的训练流程示意图；

[0053] 图4为本申请实施例提供的一种预训练阶段的流程示意图；

[0054] 图5为本申请实施例提供的一种微调训练阶段的流程示意图；

[0055] 图6为本申请实施例提供的一种构建指令向量库的原理示意图；

[0056] 图7为本申请实施例提供的一种语音指令的识别流程示意图；

[0057] 图8a和图8b为本申请一示例中提供的用户界面的示意图；

[0058] 图9为本申请实施例提供的一种数据处理装置的结构示意图；

[0059] 图10为本申请实施例所适用的一种电子设备的结构示意图。

具体实施方式

[0060] 下面结合本申请中的附图描述本申请的实施例。应理解，下面结合附图所阐述的实施方式，是用于解释本申请实施例的技术方案的示例性描述，对本申请实施例的技术方案不构成限制。

[0061] 本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本申请实施例所使用的术语“包括”以及“包含”是指相应特征可以实现为所呈现的特征、信息、数据、步骤、操作、元件和/或组件，但不排除实现为本技术领域所支持其他特征、信息、数据、步骤、操作、元件、组件和/或它们的组合等。应该理解，当我们称一个元件被“连接”或“耦接”到另一元件时，该一个元件可以直接连接或耦接到另一元件，也可以指该一个元件和另一元件通过中间元件建立连接关系。此外，这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的术语“和/或”指示该术语所限定的项目中的至少一个，例如“A和/或B”可以实现为“A”，或者实现为“B”，或者实现为“A和B”。在描述多个(两个或两个以上)项目时，如果没有明确限

定多个项目之间的关系，这多个项目之间可以是指多个项目中的一个、多个或者全部，例如，对于“参数A包括A1、A2、A3”的描述，可以实现为参数A包括A1或A2或A3，还可以实现为参数A包括参数A1、A2、A3这三项中的至少两项。

[0062] 需要说明的是，在本申请的可选实施例中，所涉及到的用户信息(如用户对应的语音数据)等相关的数据，当本申请以上实施例运用到具体产品或技术中时，需要获得用户许可或者同意，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。也就是说，本申请实施例中如果涉及到与用户有关的数据，这些数据需要经由用户授权同意、且符合国家和地区的相关法律法规和标准的情况下获取的。

[0063] 可选的，本申请实施例提供的数据处理方法，可以基于人工智能(Artificial Intelligence，AI)技术实现。比如，待处理数据的特征提取、候选标准数据的特征提取、以及训练数据集中数据的特征提取，可以通过训练好的神经网络模型实现。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。随着人工智能技术研究和进步，人工智能技术已经在多个领域广泛展开研究和应用，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。

[0064] 可选的，本申请实施例所涉及的数据处理可以基于云技术(Cloud technology)实现，比如，上述神经网络模型的训练中涉及到的数据计算、对待处理数据进行处理时涉及的数据计算可以采用云技术实现。云技术是指在广域网或局域网内将硬件、软件、网络等系列资源统一起来，实现数据的计算、储存、处理和共享的一种托管技术。云技术基于云计算商业模式应用的网络技术、信息技术、整合技术、管理平台技术、应用技术等的总称，可以组成资源池，按需所用，灵活便利。云计算技术将变成重要支撑。云计算则是指IT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。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。

[0065] 为了更好的理解和说明本申请实施例提供的方案，下面先对本申请实施例所涉及的一些相关技术用语进行说明。

[0066] 二分类交叉熵误差：即交叉熵损失，是深度学习中的一种目标函数/损失函数，用于度量预测结果分布(神经网络输出的预测结果)和真实标记(即样本标签)之间相似性，该误差是基于的样本预测结果(通常是0～1之间的概率)以及真实标记(0或1)计算得到。假设样本的预测结果为真的概率为y，真实标签y′，则对应的误差L可以表示为：$L=-y'\log(y)-(1-y')\log(1-y)$。

[0067] 语音-转译文本对：语音音频(即语音信号/语音数据)和对应的转译文本(文本数据)构成一对。

[0068] 语音-指令对：指令语音和表达该指令意图的自然语言文本，如语音“标记物品A”与文本“这里有物品A”可以是一对，语音“标记物品A”也可以与文本“标一下物品A”是一对。

[0069] MFCC(Mel-Frequency Cepstral Coefficients，梅尔倒谱系数)特征：MFCC特征是

语音处理的一种音频特征,可用于神经网络模型输入。

[0070] 音素:语音的基本声学单元,是根据语音的自然属性划分出来的语音单位,依据音节里的发音动作来分析,一个动作构成一个音素。

[0071] CNN(Convolutional Neural Networks,卷积神经网络):CNN是一种深度学习的网络结构,可以捕获输入的局部信息。

[0072] Transformer网络:一种基于注意力机制的深度学习的网络结构,可应用于文本、语音等序列输入。

[0073] CTC(Connectionist Temporal Classifcation,连续时序分类)误差:也可以称为CTC损失,是深度学习中用于让模型自动学会对齐的目标函数。

[0074] MSE(Mean Squared Error,均方误差):是深度学习中的一种损失函数,用于计算两个向量的距离。

[0075] 下面对本申请提供的多种可选实施例的技术方案以及本申请的技术方案产生的技术效果进行说明。需要指出的是,下述实施方式之间可以相互参考、借鉴或结合,对于不同实施方式中相同的术语、相似的特征以及相似的实施步骤等,不再重复描述。

[0076] 图1示出了本申请实施例提供的一种数据处理方法的流程示意图,该方法可以由任意的电子设备执行,如可以由用户终端或服务器执行,还可以由用户终端和服务器交互完成。例如,待处理数据可以用户的语音指令,用户终端可以通过执行本申请实施例提供的方法,方便、快捷的识别出用户的该语音指令的具体内容(目标标准数据,即用户意图的标准文本表达),并可以根据识别结果执行对应的操作。再例如,该方法也可以由服务器执行,服务器可以接收用户终端发送来的用户的语音指令,通过执行本申请实施例提供的方法,识别出用户的语音指令的具体内容并执行相应的操作。其中,上述用户终端包括但不限于手机、电脑、智能语音交互设备、智能家电、车载终端、可穿戴电子设备、AR/VR设备等。上述服务器可以是云服务器或物理服务器,可以是一个服务器,也可以是服务器集群。

[0077] 本申请实施例提供的方法,可以适用于任何需要根据一种模态的数据识别出与该数据相匹配的另一种模态的数据的应用场景中,其中,本申请实施例中的“模态”是指数据的形式,呈现给人们的数据样式,也就是数据的种类,比如语音数据是一种模态的数据,文本数据是另一种模态的数据。例如,本申请实施例提供的方法可以实现为应用程序的功能模块/插件,比如,应用程序可以是游戏应用,通过将本申请实施例提供的数据处理方法应用于游戏应用中,用户在玩游戏时,可以发起语音指令,通过该功能模块可以快捷的确定出用户的语音指令的识别结果(目标标准数据),游戏服务器可以根据该识别结果执行相应的操作,并可以将操作结果展示给用户。

[0078] 其中,本申请实施例提供的该方法可以适用于任何有不同模态数据匹配需求的游戏中,可以包括但不限于动作类、冒险类、模拟类、角色扮演类、休闲类等类型的游戏,例如,该方法可以用于战术竞技类游戏或竞赛类游戏中,在该类游戏中,玩家可以通过操作,在虚拟游戏场景中的游戏地图上收集各种游戏资源(如虚拟游戏道具),可选的,玩家可以通过发起语音指令进行游戏资源的收集,基于本申请实施例提供的方案,可以通过将玩家的语音指令的特征(即第一数据特征)与游戏数据库(即该应用场景中的目标数据库)中预配置的意图文本特征(即标准语音指令对应的标准文本表达的特征,也就是本申请实施例中的第二数据特征)匹配,快速、准确地确定出玩家的意图(即相匹配的标准文本表达,也就是目

标标准数据)，从而可以根据该意图执行相应的操作，完成玩家的语音指令对应的游戏资源的收集。

[0079] 下面结合图1所示的流程示意图，对本申请实施例提供的数据处理方法进行展开描述。如图1中所示，本申请实施例提供的数据处理方法可以包括以下步骤S110至步骤S140。

[0080] 步骤S110：获取待处理数据，其中，待处理数据为第一模态的数据。

[0081] 步骤S120：提取待处理数据的第一数据特征。

[0082] 其中，对于第一模态的数据具体是何种形式的数据，本申请实施例不做限定。可选的，第一模态的数据可以包括但不限于文本、语音、视频或图像中的至少一种。比如，在一些应用场景中，待处理数据可以是获取的用户的语音数据，在另一些应用场景中，待处理数据可以是用户输入的文本数据。再比如，待处理数据可以包括文本和图像两种类型的数据。

[0083] 待处理数据的第一数据特征可以是通过已经训练好的第一特征提取网络提取得到的，其中，第一特征提取网络的输入可以是待处理数据，也可以是将待处理数据进行预处理后的、符合该网络输入数据格式要求的数据。例如，可以通过预处理方式对待处理数据进行初始特征的提取，将提取的初始特征输入到上述特征提取网络中，通过该网络进一步提取到具有更好的特征表达能力的第一数据特征。

[0084] 作为一个示例，待处理数据可以是语音数据，可以先对该语音数据进行时频变换，以得到该语音数据的音频特征(如梅尔频谱特征或MFCC特征等)，可以将该音频特征作为第一特征提取网络的输入，通过该网络得到该语音数据的高层特征表示(也可以称为语音向量表示或语音表示向量)，也就是该示例中的第一数据特征。

[0085] 步骤S130：将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各第二数据特征对应的匹配结果。

[0086] 步骤S140：根据各第二数据特征对应的匹配结果，从各候选标准数据中确定出与待处理数据相匹配的目标标准数据。

[0087] 本申请实施例中，目标数据库中包括至少一个候选标准数据以及每个候选标准数据的第二数据特征，候选标准数据为第二模态的数据。

[0088] 同样的，对于第二模态的数据的具体形式本申请实施例也不做限定，可以是包括但不限于文本、语音、视频或图像中的至少一种。可以理解的是，第一模态的数据和第二模态的数据不是同一种模态的数据。比如，第一模态的数据可以是语音数据，第二模态的数据可以是文本数据。

[0089] 需要说明的是，本申请实施例中，第一模态的数据和第二模态的数据可以只包含一种类型的数据，第一模态的数据或第二模态的数据也可以包括两种或两种以上类型的数据，在第一模态的数据或第二模态的数据中的至少一个是包括两种类型的数据时，第一模态的数据和第二模态的数据为不同模态的数据，可以理解为第一模态的数据和第二模态的数据中至少存在一种类型的数据是不同的，比如，第一模态的数据可以是包括文本和图像的数据，第二模态的数据可以是语音数据。以待处理数据是包括两种类型的数据为例，待处理数据的第一数据特征可以是通过将待处理数据所包含的两种类型的数据的特征融合(如拼接或相加)得到，作为一个示例，比如，待处理数据包括语音数据和文本数据，对于待处理数据，可以分别提取该语音数据的特征和该文本数据的特征，可以通过将两部分数据的特

征进行融合得到待处理数据的数据特征。

[0090] 本申请实施例中的标准数据(上述的候选标准数据、以及下文中的第一标准数据),可以理解成基准数据或参考数据,是信息的标准表达,标准数据可以是根据应用需求预先配置好的。

[0091] 本申请实施例中,各候选标准数据的第二数据特征也可以是通过训练好的神经网络模型提取得到的,具体的,可以是通过第二特征提取网络对各个候选标准数据分别进行特征提取,得到各候选标准数据的第二数据特征。同样的,第二特征提取网络的输入可以是候选标准数据,也可以是对候选标准数据进行预处理后,再将预处理后的数据输入第二特征提取网络,得到该候选标准数据的第二数据特征。比如,候选标准数据可以是文本数据,可以通过词嵌入(Embedding)或独热编码等方式获取到文本数据的初始特征表示,将该初始特征表示输入到第二特征提取网络,得到该文本数据的高层特征表示,即第二数据特征。

[0092] 作为一可选方案,上述目标数据库中的候选标准数据可以是与标准数据库中的各第一标准数据相匹配的标准表达,第一标准数据为第一模态的数据,一个第一标准数据对应至少一个标准表达。

[0093] 一个第一标准数据以及该标准数据的标准表达,可以理解为一种信息的两种不同数据形式的标准描述/表达方式,比如,第一标准数据是语音形式的数据,标准表达是文本形式的数据,第一标准数据和对应的标准表达就可以理解为同一信息的语音表达和文本表达。比如,作为一个示例,信息是“你好”,第一标准数据是“你好”的语音数据,对应的标准表达是文本内容“你好”。

[0094] 作为一个示例,在游戏应用中可以预配置语音指令库,该语音指令库中可以存储有该游戏应用中所支持的各种各样的标准语音指令(该应用场景中的第一标准数据)对应的标准文本表达(也可以理解为语音指令的识别结果),游戏玩家可以在该游戏应用的客户端输入语音指令(待处理数据),游戏服务器可以通过执行上述步骤S120至步骤S140中,从标准语音指令对应的标准文本表达中找出与用户当前输入的语音指令相匹配的语音识别结果。

[0095] 可选的,一个第二数据特征对应的匹配结果可以是第一数据特征和该第二数据特征的匹配程度,如相似度,在得到第一数据特征和各第二数据特征的匹配结果之后,可以将匹配程度最高的第二数据特征对应的候选标准数据作为目标标准数据,也可以是按照匹配程度由高到低的顺序,将排序靠前的设定数量的匹配程度对应的候选标准数据作为目标标准数据,还可以是将匹配程度大于设定值的各第二数据特征对应的候选标准数据作为目标标准数据。

[0096] 本申请实施例提供的数据处理方法,是一种新颖的数据处理方式,采用该方法在对待处理数据进行处理时,可以无需待处理数据的数据特征进行复杂的数据识别,而是可以通过将特征匹配的方式,方便、快捷的实现了不同模态数据之间的匹配,基于该方法,由于无需对待处理数据的第一数据特征进行进一步的识别,因此,可以大大减少计算量,提高数据处理效率。进一步的,由于目标数据库存储的都是候选标准数据,通过本申请实施例的该方法所确定出的也就是与第一模态的待处理数据相匹配的第二模态的标准数据,识别准确性亦可以得到很好的保证。

[0097] 此外,在实际应用中,在需要基于待处理数据执行相应操作时,由于确定出的是与

待处理数据相匹配的标准数据，可以直接基于该标准数据执行操作，与现有一些通过对数据特征进一步识别得到对应的识别结果的方式相比，可以无需再将识别结果进一步处理为规范化的数据，可以更好的满足实际应用需求。

[0098] 本申请的可选实施例中，该数据处理方法还可以包括：

[0099] 在标准数据库存在新增第一标准数据时，获取新增第一标准数据对应的至少一个标准表达；

[0100] 提取新增第一标准数据对应的各标准表达的第二数据特征；

[0101] 将新增第一标准数据对应的各标准表达、以及各标准表达对应的第二数据特征关联存储至目标数据库中。

[0102] 在标准数据库中有新的第一标准数据时(比如，随着游戏应用的不断更新优化，游戏应用具有了更多的功能，可以支持更多的语音指令的输入)，可以通过获取这些新增标准数据对应的标准表达，并提取标准表达的第二数据特征，将标准表达(即新增的候选标准数据)和对应的第二数据特征关联存储至目标数据库中，实现对目标数据库的扩充。

[0103] 其中，在获取到待处理数据的第一数据特征之后，如果是采用现有技术中(例如，采用神经网络模型对该第一数据特征进行处理的方式来得到待处理数据对应的第二模态的匹配数据)，在有新的第一标准数据出现时，需要重新训练神经网络模型，以使得如果待处理数据是与新增数据对应的标准表达时，神经网络模型才能够支持对待处理数据的识别，方案实现复杂、成本高。而本申请实施例提供的方案，当有新增的第一标准数据出现时，只需要将新增的第一标准数据对应的标准表达以及该标准表达的第二数据特征加入目标数据库，通过该将待处理数据的第一数据特征与更新后的目标数据库中的第二数据特征的匹配，即可从候选标准数据中确定出待处理数据对应的目标标准数据，实现简单、成本低。

[0104] 本申请的可选实施例中，该数据处理方法还可以包括：根据第一数据特征，确定待处理数据的数据类型；此时，上述步骤S130中，将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，可以包括：

[0105] 在待处理数据的数据类型为指定类型时，将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配。

[0106] 在一些应用场景中，可能只需要对某个或某些指定类型的数据进行特定的处理。为了满足该应用需求，作为一可选方案，在将待处理数据的数据特征和候选标准数据的数据特征进行匹配处理前，可以预先判断待处理数据的数据类型，在待处理数据的数据类型是指定类型时，再进行匹配处理，以减少不必要的数据处理，节省计算资源。该可选方案中，目标数据库中的候选标准数据可以是指定类型的各第一标准数据对应的标准表达。在实际应用中，指定类型可以是一种类型，也可以是至少两种类型，可以根据实际需求配置。

[0107] 需要说明的，在实际实施该可选方案时，可以先执行待处理数据的数据类型的判断，在数据类型是指定类型时，再进行后续匹配处理。也可以是数据类型的判断和匹配处理都执行，之后根据匹配结果和数据类型的判别结果，再确定后续的进一步处理方式。例如，在提取得到待处理数据的第一数据特征之后，可以基于该特征确定待处理数据的数据类型(，并将第一数据特征与目标数据库中的第二数据特征进行匹配，之后根据确定出的数据类型和各第二数据特征对应的匹配结果，确定待处理数据对应的目标标准数据。可以理解的是，待处理数据的目标标准数据可能存在，有可能不存在，比如，在不满足以下条件中的任

一项时，可以确定不存在与待处理数据相匹配的目标标准数据：

[0108] 数据类型不是指定类型；各第二数据特征对应的匹配程度均小于设定值。

[0109] 本申请实施例中，根据待处理数据的第一数据特征，确定待处理数据的数据类型，也可以是通过神经网络模型实现，比如，可以通过分类模型实现。具体的，可以将待处理数据的第一数据特征输入到训练好的分类模型中，通过该模型也可以得到待处理数据的数据类型属于指定类型的概率，可以根据该概率确定待处理数据是否是指定类型。其中，分类模型可以是二分类模型，即模型对应的分类类别包括指定类型和非指定类型两种类型，模型的输出可以包括待处理数据的数据类型分别属于指定类型和非指定类型的第一概率和第二概率，可以根据第一概率和第二概率确定数据类型是否指定类型比如，第一概率大于设定概率，确定待处理数据的数据类型为指定类型。

[0110] 作为一可选方案，指定类型可以包括至少两种类型，分类模型可以是多分类模型，该模型对应的分类类别包括非指定类型、以及各个指定类型，如指定类型有两种，记为第一类型和第二类型，那么分类模型对应的分类类别可以是三类。可以通过模型预测出待处理数据分别属于非指定类型、第一类型和第二类型的概率，可以将其中最大的概率值对应的类型确定为待处理数据的数据类型。可选的，对于该方案，目标数据库可以包含多个子库，每个子库对应一种指定类型的第一标准数据(第一模态的数据)对应的标准表达(第二模态的数据)，通过分类模型可以识别出待处理数据是否是指定类型，如果是指令类型，还可以识别出具体是哪种指定类型，相应的，可以将待处理数据的第一数据特征和该指定类型对应的子库中候选标准数据的第二数据特征进行匹配，而无需将第一数据特征和各个子库中的第二数据特征都进行匹配，可以进一步减少数据处理量。

[0111] 由前文的描述可知，本申请实施例中，对于第一模态的数据(如待处理数据)，其数据特征可以是通过第一特征提取网络提取得到的；对于第二模态的数据(如各候选标准数据)，其数据特征可以是通过第二特征提取网络提取得到的；其中，第一特征提取网络和第二特征提取网络是基于训练数据集对神经网络模型进行训练得到的。

[0112] 本申请实施例中，该神经网络模型包括第一网络模型和第二网络模型，可以基于训练数据集对第一网络模型和第二网络模型进行迭代训练，将训练好的第一网络模型作为上述第一特征提取网络，将训练好的第二网络模型作为上述第二特征提取网络。对于第一网络模型和第二网络模型的模型结构本申请实施例不做限定，可以根据应用需求进行配置，比如，第一网络模型和第二网络模型都可以采用基于CNN的模型。可选的，可以根据模型要处理的数据的形式配置模型结构，比如，第一模态的数据是语音数据，第一网络模型可以采用能够很好的提取语音数据的特征的网络结构，比如Wac2vec模型，如果第二模态的数据是文本数据，第二网络模型可以采用对文本数据具有很好效果的网络结构，比如，可以采用基于Transformer网络的模型，如基于Bert(Bidirectional Encoder Representation from Transformers，基于Transformer的双向编码网络)模型的结构。

[0113] 可选的，本申请实施例中的上述包括第一网络模型和第二网络模型的神经网络模型，可以是通过以下方式训练得到的：

[0114] 获取训练数据集，其中，训练数据集包括第一训练集，第一训练集中的每个第一样本包括第一模态的第一数据、以及与该第一数据相匹配的第二模态的第二数据；

[0115] 基于上述训练数据集对初始的神经网络模型进行迭代训练，直至训练总损失值满

足预设训练结束条件，将满足训练结束条件时的第一网络模型作为第一特征提取网络，将满足训练结束条件时的第二网络模型作为第二特征提取网络；上述训练的过程可以包括下述步骤：

[0116] 将各第一数据输入到第一网络模型中，得到各第一数据的特征，将各第二数据输入到第二网络模型中，得到各第二数据的特征；

[0117] 基于各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值；其中，第一负例包括一个第一样本的第一数据和另一个第一样本的第二数据；

[0118] 若第一训练损失值不满足第一预设条件，则对第一网络模型和第二网络模型的模型参数进行调整，其中，训练总损失值满足预设训练结束条件包括第一训练损失值满足第一预设条件。

[0119] 在对上述神经网络模型进行训练时，上述第一训练集中的每个第一样本的第一数据和第二数据是相互匹配的两种模态的数据，第一样本也可以称为正例即正样本，上述第一负例(负样本)是不同的第一样本中的第一数据和第二数据，也就是不匹配的两种模态的数据，对于任一第一数据，该数据可以多个其他第二数据(除了与该第一数据相匹配的第二数据之外的第二数据)分别构成负例。在训练过程中，训练损失值是基于正样本的数据特征之间的匹配程度以及负样本的数据特征之间的匹配程度确定的。

[0120] 对于在训练过程所选择的损失函数的具体形式本申请实施例不做限定，模型训练的目的是使得相互匹配的第一数据和第二数据的特征之间的相似度尽可能大，不匹配的第一数据和第二数据的特征之间的相似度尽可能的小。

[0121] 其中，对于正样本而言，可以计算通过第一网络模型学习到的第一数据的特征和通过第二网络模型学习到的第二数据特征之间的差异程度(比如，1减去相似度，或者是两个特征之间的均方误差等)，得到对应的训练损失，对于负样本而言，一种可选方式是可以计算通过第一网络模型学习到的第一数据的特征和通过第二网络模型学习到的第二数据特征之间的匹配程度，得到对应的训练损失，通过不断的训练学习，可以使得模型学习到的正样本的数据特征之间的匹配程度越来越高(即差异越来越小)，负样本的数据特征之间的匹配程度越来越低。对于不同的损失函数，计算上述匹配程度或差异程度的计算方式也是不同的。

[0122] 本申请的可选实施例中，上述基于各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值，可以包括：

[0123] 确定各第一样本的第一数据的特征和第二数据的特征之间的差异程度，得到第一损失值；

[0124] 对于每个第一数据，确定该第一数据对应的第一相似度以及该第一数据对应的第二相似度，其中，第一相似度是该第一数据的特征和与该第一数据相匹配的第二数据的特征之间的相似度，第二相似度是该第一数据所在的第一负例中第一数据的特征和第二数据的特征之间的相似度；

[0125] 获取各第一数据对应的参考标签，参考标签包括第一相似度对应的相似度标签和第二相似度对应的相似度标签；

[0126] 基于各第一数据所对应的预测相似度和参考标签，确定第二损失值，其中，预测相似度包括第一相似度和第二相似度，第二损失值表征了各第一数据所对应的预测相似度和参考标签之间的差异；

[0127] 根据第一损失值和第二损失值，确定第一训练损失值。

[0128] 可选的，上述第一损失值可以是各正样本中第一数据的特征和第二数据的特征之间均方误差的和，也可以通过是计算各正样本中第一数据的特征和第二数据的特征之间的相似度，1减去相似度作为差异程度，将各正样本对应的差异程度之和作为第一损失值。第一损失值可以让模型学习到的正样本中两种模态的数据的特征之间尽可能接近。

[0129] 上述第二损失值也可以称为匹配误差，用于约束模型学习到的正样本中两个数据的特征之间的相似度高于负样本中两个数据的特征之间的相似度。在计算该部分损失时，上述参考标签是训练时的真实标签，也就是希望模型能够学习到的结果，具体的，对于每个第一数据，对应的真实标签中的第一相似度对应的相似度标签指的是该第一数据与其相匹配的第二数据之间的理想相似度，比如可以是1或者比较高的相似度，真实标签中的第二相似度指的第二数据与其不匹配的第二数据之间的理想相似度，比如可以是0或者比较小的相似度，参考标签可以是预配好的。基于模型输出的第一数据的特征和第二数据的特征，可以计算得到每个第一数据对应的第一相似度和各个第二相似度，可以将这些相似度构成一个相似度向量，通过计算该相似度向量与参考标签之间的差异，得到第二损失值，比如，可以将该相似度向量作为模型预测出的概率分布，将参考标签作为真实概率分布即标签，通过计算两者之间的交叉熵损失得到第二损失值。

[0130] 本申请的可选实施例中，上述将各第一数据输入到第一网络模型中，得到各第一数据的特征，可以包括：

[0131] 对于每个第一数据，通过第一网络模型对该第一数据执行以下操作，得到该第一数据的特征：

[0132] 将该第一数据划分为至少两个子数据，得到该第一数据对应的子数据系列；基于词典，提取得到子数据序列中各个子数据的特征，其中，词典包括多个数据元素，每个子数据的特征包括的特征值的个数等于词典中元素的数量，一个特征值表征了该子数据中包含词典中与该特征值的位置相对应的数据元素的概率；基于各子数据的特征，得到第一数据的特征；

[0133] 该方案中，该数据处理方法还可以包括：

[0134] 对于每个第二数据，基于词典，确定该第二数据对应于该词典的数据特征，该数据特征表征了该第二数据对应于词典中各个数据元素的概率；

[0135] 相应的，上述确定第一训练损失值可以包括：

[0136] 基于各第一样本中的第一数据的各个子数据的特征与第二数据对应于词典的数据特征之间的匹配程度、各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值。

[0137] 可以看出，该可选方案中，第一训练损失值还增加了各第一样本中的第一数据的各个子数据的特征与第二数据对应于上述词典的数据特征之间的匹配程度(可以称为第三损失部分)对应的损失(可以称为第三损失值)，基于该损失，可以使得基于第一网络模型学

习到的第一数据中各个子数据的特征，能够预测得到与该第一数据相匹配的第二数据的概率最大化，也就是说，第三损失值是为了能够约束第一网络模型，让该模型学习到的第一数据中各个子数据的特征能够预测得到第二数据。

[0138] 可选的，第三损失部分可以采用CTC误差(也可以叫CTC损失)，采用该误差可以让模型自动学会不同模态的数据之间的对齐。本申请实施例中，上述词典中的数据元素是能够用于表示第一数据的各子数据和第二数据的数据单元，对于数据元素的形式可以根据需求配置，可选的，数据元素可以包括但不限于拼音或音素。以音素为例，词典中数据元素包括各个音素和一个空白符(CTC损失中为了实现自动数据之间的自动对其增加的一个伪标识，也称为blank)。对于第一数据的各个子数据，子数据的特征的维度(也就是特征向量的长度)等于词典中数据元素的数量，词典中各个数据元素的位置是固定的，子数据的特征表征的是该子数据中包含每个位置的数据元素的概率，作为一个示意性的说明，假设词典中有a、b、c三个元素，一个子数据的特征的长度为3，可以表示为(p1，p2，p3)，p1、p2和p3分别表示在第一个位置出现a的概率为p1、在第二个位置出现b的概率为p2、在第三个位置出现c的概率为p3。对于第二数据而言，其对应于词典的数据特征是表征该数据特征表征了该第二数据对应于词典中各个数据元素的概率。在计算每个正样本对应的第三损失值时，可以基于第一数据的各个子数据的特征序列(也就是各个特征值组成的特征向量)，确定可以根据这些子数据的特征序列得到第二数据对应于词典的数据特征的概率，基于第三损失值的约束，使得该概率达到最大化，可以让第一网络模型学习到的第一数据的各子数据的特征中能够包含第二数据的语义信息。

[0139] 可选的，第一数据可以为语音数据，第二数据可以为文本数据，词典中的数据元素可以为音素。第二数据对应于词典的数据特征，可以是第二数据对应的音素序列，也就是构成该第二数据的各个音素组成的序列，在对语音数据进行特征提取时，可以先基于词典，对该语音数据的各个语音帧(即子数据)先进行特征提取，得到每个语音帧对应的特征表示，在计算CTC损失时，文本数据对应的音素序列作为标签，根据各个语音帧的特征表示和该标签，计算得到CTC损失，该损失的值表征了根据各语音帧的特征表示预测得到文本数据的音素序列的概率，概率越大，损失的值越小。

[0140] 本申请的可选实施例中，上述候选标准数据可以为指定类型的第一标准数据对应的第二模态的标准表达；上述初始的神经网络模型还包括分类模型；此时，练数据集还包括第二训练集，第二训练集中的每个第二样本包括第一模态的第三数据、与该第三数据相匹配的第二模态的第四数据、以及该第三数据的类型标签，其中，第二训练数据集中的第三数据包括指定类型的第三数据和非指定类型的第三数据；在得到第一训练损失值满足第一预设条件的神经网络模型之后，模型的训练的过程还可以包括：

[0141] 基于第二训练集继续对神经网络模型重复执行训练操作，直至第二训练损失值满足第二预设条件，其中，训练总损失值满足预设训练结束条件还包括所述第二训练损失值满足第二预设条件；该训练操作可以包括：

[0142] 将各第三数据输入到第一网络模型中，得到各第三数据的特征，将各第四数据输入到第二网络模型中，得到各第四数据的特征，将各第三数据的特征的输入至分类模型中，得到各第三数据对应的预测类型；

[0143] 基于各第二样本中的第三数据的特征与第四数据的特征的匹配程度、各第二负例

中的第三数据的特征和第四数据的特征的匹配程度、以及各第三数据的类型标签和预测类型之间的匹配程度，确定第二训练损失值；

[0144] 若第二训练损失值不满足第二预设条件，则对神经网络模型的模型参数进行调整。

[0145] 由前文的描述可知，在一些应用场景中，需要对待处理数据的数据类型进行判别，可以在数据类型是指定类型再进行进一步处理。为了满足该应用需求，本申请的该可选实施例中，上述神经网络模型中除了包括第一网络模型和第二网络模型之外，还可以包括分类模型，该分类模型与第一网络模型级联，用于根据第一网络模型输出的特征判别输入至第一网络模型的数据的类型。该可选实施例中，前文中基于第一训练集对神经网络模型进行训练的过程可以称为预训练，通过预训练，可以得到基本满足应用需求的第一网络模型和第二网络模型，在通过预训练得到满足第一预设条件的神经网络模型(为描述方便，将该模型称为中间模型)之后，可以基于第二训练数据集对该中间模型继续进行微调训练，以得到能够更加更好的满足特定任务需求的模型。

[0146] 在微调训练过程中，可以根据各第二样本中的第三数据的特征与第四数据的特征的匹配程度、以及各第二负例中的第三数据的特征和第四数据的特征的匹配程度计算一部分训练损失(匹配损失)，可以根据各第三数据的类型标签和预测标签计算一部分训练损失(分类损失)，基于这两部的训练损失来约束模型的进一步训练。其中，根据各第二样本中的第三数据的特征与第四数据的特征的匹配程度、以及各第二负例中的第三数据的特征和第四数据的特征的匹配程度计算损失的方式，可以采用前文中计算匹配损失(即第二损失值)的方式，当然，也可以是前文中计算第一损失值和第二损失值的方案。

[0147] 对于分类损失，该部分的损失值表征的是通过分类模型预测得到的第三数据的类型和第三数据的真实类型即类型标签的相似性，可选的，第三数据的类型标签可以是1或0，比如，1表示第三数据是指定类型的数据，0表示第三数据不是指定类型的数据，分类模型的输出可以包括第三数据是指定类型的第一概率和第三数据不是指定类型的第二概率，可以根据各第三数据的类型标签分类模型输出的两个概率，计算分类模型对应的训练损失部分，可选的，该损失部分可以采用二分类交叉熵误差计算得到，误差值越小代表预测出的类型和真实类型越接近。

[0148] 在通过该可选实施例得到满足训练结束条件的神经网络模型之后，在应用时，可以通过训练好的分类模型来识别待处理数据的数据类型，具体的，可以将待处理数据输入至训练好的第一网络模型(即第一特征提取网络)中，得到待处理数据的第一数据特征，将该第一数据特征输入到训好的上述分类模型中，得到待处理数据属于指定类型数据的第一概率和不属于指定类型数据的第二概率，根据第一概率和第二概率可以确定待处理数据是不是指定类型的数据。比如，待处理数据是语音数据，指令类型是语音指令，也就是说指定类型的语音数据是指令型的语音，如果通过分类模型确定出待处理数据是指令型的语音数据，可以将该其特征和目标数据库中的各候选标准数据(如文本)的特征进行匹配，找到与该语音数据相匹配的文本表达，也可以理解为语音数据的识别结果。

[0149] 本申请实施例提供的数据处理方案，提出一种基于跨模态检索的数据匹配方法，该方法可以只利用待处理数据的数据特征，而不需要对该数据特征进行进一步深度识别，就可以从目标数据库中快速的找出与该待处理数据相匹配的另一种模态的数据即目标标

准数据。相比于现有技术，本申请的该方案可以大大有效减少数据计算量，且准确率与现有技术相比也有了明显了提升。

[0150] 本申请实施例提供的方法，可以适用于任何需要进行跨模态数据间的处理场景中，比如，该方法可以应用于AI语音助手的指令识别场景中，可以准确、快速的识别出用户的语义指令；该方法还可以应用于AI机器人的语音问答中，通过该方法，可以找到与用户输入的语音相匹配的文本表达，从而可以将该文本表达对应的答案信息提供给用户；该方法还可以应用于跨模态的数据检索场景中，比如，可以应用于搜索引擎或各种应用程序中，可以基于用户输入的文本数据找到相匹配的音频数据，比如，可以根据用户输入的搜索文本找到对应的音乐提供给用户。另外，本申请实施例提供的上述第一特征提取网络和第二特征提取网络还可以应用于各种有需要提取数据特征的场景中，可以提取得到具有更好的语义表达能力的数据特征。

[0151] 在实际实施时，本申请实施例中待处理数据和候选标准数据都可以是包括至少一种类型的数据，如第一模态的数据为语音，第二模态的数据是文本。可以理解的是，在待处理数据或第二模态的数据可以是包括两种类型的数据时，在对上述第一特征提取网络和第二特征提取网络进行训练时，训练数据集中的第一模态的数据(第一数据和第三数据)和第二模态的数据(第二数据和第四数据)，应至少包括分别与待处理数据和候选标准数据相对应的类型的数据，比如，待处理数据可以是包括第一类型的数据和第二类型数据的数据，候选标准数据为第三类型的数据，那么训练数据集中的至少部分第一数据和第三数据也应包括第一类型的数据和第二类型的数据，这部分第一数据对应的第二数据以及这部分第三数据对应的第四数据也应是第三类型的数据，即对特征提取网络进行训练时所使用的训练数据集中的样本数据的类型应与网络训练好之后所处理的数据的类型是对应的。

[0152] 为了更好的理解本申请实施例提供的方法以及该方法的实用价值，下面结合具体的场景实施例，对本申请实施例提供的方法进行说明。

[0153] 该场景实施例对应的应用场景为游戏场景，可以将本申请实施例提供的方法，应用于游戏应用中的AI语音助手中，A语音助手可以对用户输入的语音指令进行识别。在游戏场景中，用户在玩游戏时，可以通过语音与AI语音助手交互，比如，用户在其用户终端中玩游戏时说了“标记P城”，其目的是让AI语音助手在游戏虚拟场景的地图的地点“P城”作出标记，语音指令“标记物品A”则是让AI语音助手对虚拟游戏场景中的“物品A”进行标记。

[0154] 图2示出了本申请该场景实施例中适用的一种数据处理系统的结构示意图，如图2中所示，该数据处理系统可以包括用户终端10、游戏服务器20和训练服务器30，用户终端10可以通过网络与游戏服务器20通信连接，用户终端10可以是任一游戏玩家的用户终端，游戏服务器20用于为玩家提供游戏服务，其中，游戏应用的类型本申请实施例不做限定，可以是需要用户下载安装的游戏应用，也可以是云游戏应用，还可以是小程序中的游戏应用。训练服务器30可以通过网络与游戏服务器20通信连接，训练服务器30可以用于执行神经网络模型的训练操作，并将训练好的神经网络模型提供给游戏服务器20中。

[0155] 其中，上述AI语音助手可以是部署于游戏服务器20中，也可以是部署于用户终端10，可选的，为了减少用户终端10的计算资源，下面以AI语音助手部署在游戏服务器20一侧为例进行说明。

[0156] 本应用场景中，待处理数据为用户的语音数据，第一标准数据为预配置的标准语

音指令，也就是游戏应用所支持的语音指令，第一标准数据对应的标准表达为文本表达(也就是候选标准数据)。下面结合图2中所示的数据处理系统，对游戏场景中本申请提供的方法的一种可选实施流程进行说明。该实施例中的数据处理流程可以包括如下步骤1至步骤5：

[0157] 步骤S1：训练神经网络模型。

[0158] 该步骤可以由训练服务器30执行。图3中示出了本场景中神经网络模型的训练原理示意图。如图3中所示，训练过程可以包括预训练和微调训练两个阶段，图4示出了预训练阶段的原理示意图，图5示出了微调训练阶段的原理示意图。

[0159] 如图3中所示，本场景实施例中的神经网络模型包括语音编码模块(第一网络模型)、文本编码模块(第二网络模型)和分类模块。语音数据通过语音编码模块可以得到对应的语音表示向量(即语音数据的特征，也可以称为向量表示)，文本数据通过文本编码模块可以得到对应的文本表示向量(即文本数据的特征)。对于该模块的具体网络结构本申请不做限定，可以根据实际需要配置。

[0160] 可选的，语音编码模块可以采用基于Wav2vec2模型的结构，比如可以在Wav2vec2模型(图4中所示的Wav2vec2)之后接一个池化模块(图4中所示的池化操作)作为语音编码模块，Wav2vec2模型由多层CNN和多层Transformer组成。采用该语音编码模块进行特征提取时，语音数据首先通过Wav2vec2模型编码得到向量序列，序列的每个值是一个向量(可以理解为一个语音段的特征向量)，然后可以通过池化操作计算向量序列的平均值得到一个向量，这个向量即语音表示向量。对于文本编码模块，可选的，该模块可以采用基于Bert模型的结构，该模型由多层Transformer组成。下面结合图3至图5对神经网络模型的预训练和微调训练这两个阶段的训练过程进行说明。

[0161] 预训练过程

[0162] 预训练是基于第一训练集进行的，第一训练集中包括大量的语音-转译文本数据(即第一样本，包括语音数据以及该语音数据相匹配的文本数据)。为了描述方便，下面将训练数据集中的语音数据(图3中的第一数据)称为样本语音，语音数据对应的文本数据(图3中的第二数据)称为转译文本，对预训练过程进行说明。

[0163] 如图3和图4中所示，一次训练过程可以包括：将各样本语音输入至语音编码模块，通过语音编码模块得到各样本语音的语音表示向量，将各转译文本输入至文本编码模块得到各转译文本的文本表示向量，之后，可以基于各样本语音的语音表示向量和各转译文本的文本表示向量计算预训练阶段的训练损失(第一训练损失值)，也就是图4中所示的总体误差，如果此次训练的总体误差满足第一预设条件，则预训练阶段可以结束，得到中间模型，如果总体误差不满足第一预设条件，则对语音编码模块和文本编码模块的模型参数进行调整(图4中所示的更新参数)，并重复上述训练过程，直至总体误差满足第一预设条件。

[0164] 如图4中所示，本场景实施例中预训练采用了多种训练目标，即多种训练误差，也就是多种训练损失，具体包括CTC误差(第三损失部分)、匹配误差(第一损失部分)和蒸馏误差(第二损失部分)，预训练阶段的第一训练损失值是这三部分的误差之和，这三部分误差的含义如下：

[0165] (1)CTC误差：该误差是基于语音编码模块的中间向量(各子数据的特征)和标签(转译文本对应于词典的数据特征)计算得到的，可以表征一个正样本中样本语音和转译文

本之间的相似度，标签是根据转译文本生成的，是对应文本的汉语拼音序列，使用拼音而不是字是为了减小词典大小，即词典中数据元素的数量，也就是标签的长度。Wav2vec2模型输出的向量序列是语音数据的各个语音段的特征，传入CTC误差，和标签一起作为计算CTC误差的输入，一个正样本对应的CTC误差L_{ctc}可以表示如下：

[0166] $$\mathcal{L}_{ctc} = -\log p_{ctc}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{\pi} \in \Phi_{\boldsymbol{x},\boldsymbol{y}}} p(\boldsymbol{\pi}|c_1, ..., c_T)$$

[0167] 在本实施例中，y表示标签，x表示样本语音，c_1，c_2，…，c_T表示样本语音经过语音编码模块的Wav2vec2模型编码后输出的各语音段(语音帧)的表示向量，π表示与样本语音x对应的合法序列，可以理解为根据样本语音的各语音段的表示向量可以得到标签y(即拼音序列)的路径，L_{ctc}则表征了可以根据样本语音的各语音段的表示向量可以得到y的所有路径对应的概率，也就是可以根据样本语音的各语音段的表示向量得到标签y的概率。

[0168] 对于一个语音-转译文本对，该误差代表了可以根据Wav2vec2模型输出的样本语义的各语音段对应于词典的特征向量预测得到标签y的可能性。在训练过程中，CTC误差的目的是令Wav2vec2模型输出的向量序列能预测转译文本的拼音，令这些向量包含语义信息。

[0169] (2)匹配误差：样本语音和转译文本对应的向量表示作为该误差的输入，匹配误差的目标是令语音-转译文本(正例)的两个向量表示的相似度高于语音-负例转译文本(负例)的两个向量表示的相似度，可选的，相似度的计算可以采用余弦相似度。训练时，一次性计算一批语音-转译文本数据，批次里的每个语音-转译文本对属于正例，而批次内其他语音、转译文本两两构成负例，即语音-负例转译文本。

[0170] 可选的，匹配误差可以采用多分类交叉熵损失，具体的，在通过模型得到各样本语音的语音向量表示和各转译文本的文本向量表示之后，对于每个样本语音，可以该语音所属的正例的两个向量表示之间的第一相似度，以及该语音所属的各负例的两个向量表示之间的第二相似度，也就是该语音的语音表示向量和其他各个文本(各转译文本中除了与该语音相匹配的转移文本之外的文本)的文本表示向量之间的相似度，作为一个示意性说明，假设一批数据中有10个语音-转移文本对，那么对于每个样本语音，第二相似度是9个，第一相似度是一个，可以将这10个相似度作为预测结果分布，例如，将第一相似度作为分布的第一个值，预测结果分布可以表示为一个分布向量[p1,p2,…,p10]，p1表示第一相似度，其他9个为第二相似度，该预测结果分布对应的真实分布(真实标签)可以表示为[1,0,…,0]，1表示样本语音与其相匹配的转移文本之间的相似度的真实标签，0表示样本语音与其他文本之间的相似度的真实标签，将真实分布作为真实标签，将预测结果分布作为模型的预测分布，通过多分类交叉熵损失，可以计算得到各样本语音对应的匹配误差。

[0171] 在训练过程中，通过匹配误差的约束，可以使得模型学习到的正例的向量表示之间的相似度高于负例的向量表示之间的相似度。

[0172] (3)蒸馏误差：正例对应的向量表示作为该误差的输入，计算两个向量的均方误差(MSE)，该误差的目的是让语音-转译文本对的两个向量表示距离接近。

[0173] 通过计算上述三个误差，把这三个误差值取平均或求和得到总体误差，如果总体误差不满足第一预设条件，则更新语音编码模块和文本编码模块的模型参数，并继续重复

训练。

[0174] 在得到满足第一预设条件的中间模型之后，可以基于第二训练集对中间模型的各部分继续进行训练，即微调训练。

[0175] 微调训练过程

[0176] 图5示出了微调训练过程的原理示意图，如图3和图5中所示，在微调训练阶段，除了继续对语音编码模块和文本编码模块进行训练外，还需要对分类模型(图5中所示的拒识分类模块)进行训练，通过微调训练，得到满足第二预设条件的神经网络模型，可以将此时的语音编码模块和分类模块部署于游戏服务器20的AI语音助手中，用于对游戏玩家输入的语音数据的类别进行识别，以及提取该语音数据的语音表示向量，可选的，文本编码模块同样也可以部署于游戏服务器，用于提取标准语音指令对应的标准文本表达的文本表示向量，或者，文本编码模块也可以部署于其他设备中，由该设备提取各标准语音指令对应的标准文本表达的文本表示向量之后，提供给游戏服务器使用。对于分类模块的结构本申请也不做限定，可选的，拒识分类模块可以由一层全连接层网络组成。下面结合图3和图5对微调训练的过程进行说明。

[0177] 在微调训练阶段，因为拒识分类模块是判断输入到语音编码模块的样本语音是否是指令，因此，第二训练集中的第二样本除了包括样本语音和该样本语音对应的转译文本之外，还包括样本语音的类型标签，该阶段，各第二样本中除了包括指令语音-指令语音的转移文本组成的样本之外，还包括非指令语音-非指令语音的转译文本组成的样本。比如，类型标签为标签是1表示是指令语音，0表示不是指令语音。

[0178] 如图5中所示，在微调训练阶段，同样的，将各样本语音(指令语音和非指令语音)通过语音编码模块分别得到各样本语音的语音表示向量，各样本语音的转移文本通过文本编码模块得到对应的文本表示向量。各样本语音的语音表示向量通过拒识分类模块得到样本语音是指令语音的概率。可选的，该阶段的目标函数(即损失函数)可以有两个，拒识分类误差和匹配误差，其中，拒识分类误差可以采用二分类交叉熵误差，该误差计算的是通过分类模型预测出的样本语音是指令语音的概率和样本语音的类型标签之间的差异，匹配误差可以采用与预训练阶段相同的匹配误差。之后，可以将两个误差取平均值或求和作为第二训练损失值(图5中的总体误差)，如果总体误差满足第二预设条件，得到训练好的模型，如果总体误差不满足第二预设条件，则对语音编码模块、文本编码模块和分类模型的模型参数进行更新，并重复上述微调训练过程。

[0179] 步骤S2：构建目标数据库。

[0180] 在模型应用和对待处理数据进行前，需要先构建目标数据库，本应用场景中的目标数据库包括如图2中所示的指令库和指令向量库。其中，指令库中存储的是各标准语音指令对应的文本形式的标准表达(候选标准数据)，指令向量库中存储的是指令库中的各个标准表达的文本向量表示(第二数据特征)。指令库构造可以是根据指令意图类别构造的，指令库包含多个具体意图(对应标准语音指令，每条标准语音指令可以理解为用户的一个意图)，每个意图可作为一个类别并赋予类别id，同时对每个意图可以构造一条或多条表达该意图的自然语言文本作为该意图的标准表达。比如，语音指令“标记物品A”可作为一个类别，其对应的标准表达可以有“标记物品A”、“这里有物品A”，“标一下物品A”等多种标准表达。

[0181] 对于指令向量库,其中的文本向量表示可以是通过训练好的文本编码模块提取得到的。如图6中所示,对于指令库中的每个标准表达,可以将其输入到文本编码模块中,通过该模块得到对应的文本向量表示,将各意图对应的标准表达的文本向量表示存储到指令向量库中。

[0182] 基于本申请实施例提供的方案,可以很方便、快捷的实现对指令数据库的更新扩展,例如,当有新的语音指令加入时,只需要将新的语音指令对应的文本表达加入到指令库中,并通过文本编码模块提取对应的文本向量表示计入到指令向量库中即可,无需重新训练模型,可以很好的满足目标数据库中指令类别(一个标准语音指令可以当作一个类别)不固定时的应用需求。

[0183] 可选的,在微调训练阶段,第二训练集中的指令语音-指令语音的转译文本对可以包括由指令库中的标准语音指令以及标准语音指令对应的标准表达构成的文本对。

[0184] 步骤S3:对待处理数据进行处理。

[0185] 在应用阶段,对于用户的每条语音输入(即语音指令,也就是本应用场景中的待处理数据),可以采用图7中所示的处理流程处理,具体的,可以用训练好的语音编码模块得到每条语音输入的语音表示向量(即第一数据特征),将该向量表示作为查询向量,计算查询向量和指令向量库的每条向量表示的相似度(即匹配分数),可以选取最大匹配分数以及该匹配分数对应的文本向量表示(图7中所示的指令和匹配分数),另外,将语音表示向量输入到拒识分类模块通过拒识分类模块得到拒识分数(如语音输入是语音指令的概率),之后可以依据匹配分数和拒识分数,由规则判断模块根据预配置的判断规则,判断最大相似度对应的文本向量表示所对应的标准表达是否是与该语音输入相匹配的目标标准数据。例如,判断规则是匹配分数大于第一阈值且拒识分数大于第二阈值,如果满足该条件,则将最大相似度对应的标准表达确定为目标标准数据,也就是说,认为用户此时输入的语音是该标准表达对应的语音指令(图7中最后输出的指令),可以根据用户的语音输入,执行该标准表达对应的标准语音指令对应的动作,比如,标准表达为“标记物品A”,游戏服务器则可以执行对应的标记动作,并将标记结果通过用户终端的用户界面(即游戏应用的交互界面)展示给玩家。

[0186] 作为一个示例,图8a和图8b中示出了一种游戏场景的用户界面的示意图,玩家在玩游戏的过程中,可以通过发起语音指令的方式或者是手动操作的方式进行游戏操作,具体的,玩家可以通过在其终端设备上或者终端设备外接的输入控制设备对其玩家角色进行控制操作,如图8a中,在游戏过程,玩家可以点击“语音助手”控件(图8a中的控件81)打开或者关闭AI语音助手,在AI语音助手打开的状态下可以发起语音指令。在该游戏场景中,玩家如果想要对场景中的虚拟物品A进行标记,可以控制其玩家角色(可以通过角色的指定虚拟道具)对该物品A进行瞄准,在瞄准之后可以发出“标记物品A”的语音指令,或者是通过点击“标记”控件(图8a中的控件81)对物品A进行标记,在采用语音方式时,AI语音助手在接收到“标记物品A”的语音指令后,可以通过执行本申请实施所提供的数据处理方法(如上述步骤S120至步骤140所对应的任一可选实施方式),或者将语音指令发送给游戏服务器,由游戏服务器通过执行该方法,确定出用户的真实意图(即目标标准数据)是需要对“物品A”进行标记,AI语音助手则可以该真实意图,对游戏场景中的该物品A进行标记。如图8b所示,在对物品A进行标记之后,可以显示对应的标记信息,该标记信息包括但不限于物品A的属性信

息(如类别等)、标记提示信息(如图8b中的“abc标记了物品A”以及悬浮在物品A上方的标记标志83,abc为玩家在游戏中的名称即玩家昵称)以及玩家角色在游戏场景中当前距离该物品的距离(图中所示的5米)等。

[0187] 在实际应用中,如果玩家当前是在组队游戏中,玩家标记了物品A之后,该玩家所在队伍中的其他玩家也可以在其用户界面看到相应的提示信息,如“abc标记了物品A”、其他玩家的玩家角色在游戏场景中与该物品A的相对位置信息等。

[0188] 作为一可选方案,为了加快检索效率,在通过计算查询向量和指令向量库的向量表示的相似度,确定匹配分数最高的标准表达时,可以采用Faiss检索方式(一种快速检索方法),相应的,在通过文本编码模块得到各标准表达的文本向量表示之后,可以采用Faiss检索方式中的特征向量库构建方式构建指令向量库,以提高检索效率。

[0189] 可见,采用本申请实施例提供的方法,在对语音数据进行识别时,可以无需语音识别模型,可以基于语音数据的语音表示向量和文本数据的文本表示向量之间的相似度,快速、准确的识别出与该语音数据相匹配的文本表达,该文本表达对应的标准语音指令就可以作为该语音数据的真实意图,根据该意图执行相应操作。采用该方法可以有效降低计算成本,提高处理效率。另外,通过语音检索文本的跨模态检索方式,可以很好的解决多指令类别情况下识别准确率不佳的问题,且能够满足指令类别的实际应用需求,当有新的指令类别时,不需要重新训练模型,只需在检索库(即指令向量库)加入新的指令文本。另外,通过预训练加微调训练两阶段的训练方式,结合多目标优化方案,让语音编码模块能够很好的学习到语音的语义信息,可以提高识别的准确率。

[0190] 需要说明的是,在实际应用中,除了可以采用本申请实施例提供的上述分阶段训练的方式之外,还可以采用将两阶段合并或两个阶段交替式进行训练的方式。

[0191] 为了验证本申请实施例提供的方法的效果,在游戏场景中,对本申请实施例提供的方法和现有方案进行了比对测试。在测试时,采用了由人工标注的1469条数据,现有方案采用了自动语音识别(ASR,Automatic Speech Recognition)加自然语音理解(NLU,Natural Language Processing)的方法,并采用了准确率、召回率和误召率作为效果评估指标,误召率是指对于非指令语音,模型给出了指令识别结果的比例。准确率和召回率越高越好,误召率越低越好。下表1中示出了本申请实施例的方案和现有方案的测试结果,需要说明的是,在测试时,测试数据采用的是筛选过的数据,非指令语音数据都很像语音指令,所以两种方案的误召率都比较高。

[0192] 表1在人工标注的数据效果对比

[0193]

模型	准确率	召回率	误召率
现有方案	38.02%	61.97%	64.20%
本申请	42.86%	76.40%	60.34%

[0194] 通过测试结果可以看出,相比于现有方案,本申请实施例提供的神经网络模型对应的准确率和召回率相比于现有方案都有明显较大提升,误召率明显下降。

[0195] 基于与本申请实施例提供的方法相同的原理,本申请实施例还提供了一种数据处理装置,如图9中所示,该数据处理装置100可以包括待处理数据获取模块110、特征获取模块120和数据识别模块130。

[0196] 待处理数据获取模块110,用于获取待处理数据,待处理数据为第一模态的数据;

[0197] 特征获取模块120，用于提取待处理数据的第一数据特征；

[0198] 数据识别模块130，用于将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各第二数据特征对应的匹配结果，以及根据各第二数据特征对应的匹配结果，从各候选数据中确定出与待处理数据相匹配的目标数据；

[0199] 其中，目标数据库中包括至少一个候选标准数据以及每个候选标准数据的第二数据特征，候选标准数据为第二模态的数据。

[0200] 可选的，数据识别模块还用于：根据第一数据特征，确定待处理数据的数据类型；相应的，数据识别模块在将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配时用于：

[0201] 在待处理数据的数据类型为指定类型时，将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配。

[0202] 可选的，第一模态的数据和第二模态的数据为不同模态的数据，第一模态的数据包括文本、语音、视频或图像中的至少一种，第二模态的数据包括文本、语音、视频或图像中的至少一种。

[0203] 可选的，候选标准数据是与标准数据库中的第一标准数据相匹配的标准表达，第一标准数据为第一模态的数据，一个第一标准数据对应至少一个标准表达。

[0204] 可选的，特征获取模块还用于：在标准数据库存在新增第一标准数据时，获取新增第一标准数据对应的至少一个标准表达；提取新增第一标准数据对应的各标准表达的第二数据特征；将新增第一标准数据对应的各标准表达、以及各标准表达对应的第二数据特征关联存储至目标数据库中。

[0205] 可选的，第一数据特征是通过第一特征提取网络提取得到的；候选标准数据的第二数据特征是通过第二特征提取网络提取得到的；其中，第一特征提取网络和第二特征提取网络是由模型训练模块通过以下方式训练得到的：

[0206] 获取训练数据集，训练数据集包括第一训练集，第一训练集中的每个第一样本包括第一模态的第一数据、以及与该第一数据相匹配的第二模态的第二数据；

[0207] 基于训练数据集对初始的神经网络模型进行迭代训练，直至训练总损失值满足预设训练结束条件，其中，神经网络模型包括第一网络模型和第二网络模型，将满足训练结束条件时的第一网络模型作为第一特征提取网络，将满足训练结束条件时的第二网络模型作为第二特征提取网络；训练的过程包括：

[0208] 将各第一数据输入到第一网络模型中，得到各第一数据的特征，将各第二数据输入到第二网络模型中，得到各第二数据的特征；

[0209] 基于各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值；其中，第一负例包括一个第一样本的第一数据和另一个第一样本的第二数据；

[0210] 若第一训练损失值不满足第一预设条件，则对第一网络模型和第二网络模型的模型参数进行调整，训练总损失值满足预设训练结束条件，包括第一训练损失值满足第一预设条件。

[0211] 可选的，模型训练模块在将各第一数据输入到第一网络模型中，得到各第一数据的特征时用于执行操作：

[0212] 对于每个第一数据，通过第一网络模型对该第一数据执行以下操作，得到该第一数据的特征：将该第一数据划分为至少两个子数据，得到该第一数据对应的子数据系列；基于词典，提取得到子数据序列中各个子数据的特征，其中，词典包括多个数据元素，每个子数据的特征包括的特征值的个数等于词典中元素的数量，一个特征值表征了该子数据中包含词典中与该特征值的位置相对应的数据元素的概率；基于各子数据的特征，得到第一数据的特征；

[0213] 模型训练模块还可以用于：对于每个第二数据，基于词典，确定该第二数据对应于词典的数据特征，该数据特征表征了该第二数据对应于词典中各个数据元素的概率；

[0214] 模型训练模块在确定第一训练损失值时可以用于：基于各第一样本中的第一数据的各个子数据的特征与第二数据对应于词典的数据特征之间的匹配程度、各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值。

[0215] 可选的，模型训练模块在基于各第一样本中的第一数据的特征与第二数据的特征的匹配程度、以及各第一负例中的第一数据的特征和第二数据的特征的匹配程度，确定第一训练损失值时，可以用于：

[0216] 确定各第一样本的第一数据的特征和第二数据的特征之间的差异程度，得到第一损失值；

[0217] 对于每个第一数据，确定该第一数据对应的第一相似度以及该第一数据对应的第二相似度，其中，第一相似度是该第一数据的特征和与该第一数据相匹配的第二数据的特征之间的相似度，第二相似度是该第一数据所在的第一负例中第一数据和第二数据的相似度；

[0218] 获取各第一数据对应的参考标签，参考标签包括第一相似度对应的相似度标签和第二相似度对应的相似度标签；

[0219] 基于各第一数据所对应的预测相似度和参考标签，确定第二损失值，其中，预测相似度包括第一相似度和第二相似度，第二损失值表征了各第一数据所对应的预测相似度和参考标签之间的差异；

[0220] 根据第一损失值和第二损失值，确定第一训练损失值。

[0221] 可选的，候选标准数据为指定类型的第一标准数据对应的第二模态的标准表达；初始的神经网络模型还包括分类模型；训练数据集还包括第二训练集，第二训练集中的每个第二样本包括第一模态的第三数据、与该第三数据相匹配的第二模态的第四数据、以及该第三数据的类型标签，其中，第二训练数据集中的第三数据包括指定类型的第三数据和非指定类型的第三数据；在得到第一训练损失值满足第一预设条件的神经网络模型之后，模型训练模块还用于执行以下训练过程：

[0222] 基于第二训练集继续对神经网络模型重复执行训练操作，直至第二训练损失值满足第二预设条件，其中，训练总损失值满足预设训练结束条件还包括第二训练损失值满足第二预设条件；上述训练操作包括：

[0223] 将各第三数据输入到第一网络模型中，得到各第三数据的特征，将各第四数据输入到第二网络模型中，得到各第四数据的特征，将各第三数据的特征的输入至分类模型中，得到各第三数据对应的预测类型；

[0224] 基于各第二样本中的第三数据的特征与第四数据的特征的匹配程度、各第二负例中的第三数据的特征和第四数据的特征的匹配程度、以及各第三数据的类型标签和预测类型之间的匹配程度，确定第二训练损失值；

[0225] 若第二训练损失值不满足第二预设条件，则对神经网络模型的模型参数进行调整。

[0226] 可选的，第一模态的数据为语音，第二模态的数据为文本，数据元素为音素。

[0227] 可选的，指定类型为指令型语音。

[0228] 本申请实施例的装置可执行本申请实施例所提供的方法，其实现原理相类似，本申请各实施例的装置中的各模块所执行的动作是与本申请各实施例的方法中的步骤相对应的，对于装置的各模块的详细功能描述及有益效果具体可以参见前文中所示的对应方法中的描述，此处不再赘述。

[0229] 本申请实施例中还提供了一种电子设备，该电子设备包括存储器、处理器及存储在存储器上的计算机程序，该处理器执行上述计算机程序以实现本申请任一可选实施例中提供的方法的步骤。

[0230] 图10示出了本申请实施例所适用的一种电子设备的结构示意图，如图10所示，该电子设备4000包括处理器4001和存储器4003。其中，处理器4001和存储器4003相连，如通过总线4002相连。可选地，电子设备4000还可以包括收发器4004，收发器4004可以用于该电子设备与其他电子设备之间的数据交互，如数据的发送和/或数据的接收等。需要说明的是，实际应用中收发器4004不限于一个，该电子设备4000的结构并不构成对本申请实施例的限定。

[0231] 处理器4001可以是CPU(Central Processing Unit，中央处理器)，通用处理器，DSP(Digital Signal Processor，数据信号处理器)，ASIC(Application Specific Integrated Circuit，专用集成电路)，FPGA(Field Programmable Gate Array，现场可编程门阵列)或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框，模块和电路。处理器4001也可以是实现计算功能的组合，例如包含一个或多个微处理器组合，DSP和微处理器的组合等。

[0232] 总线4002可包括一通路，在上述组件之间传送信息。总线4002可以是PCI(Peripheral Component Interconnect，外设部件互连标准)总线或EISA(Extended Industry Standard Architecture，扩展工业标准结构)总线等。总线4002可以分为地址总线、数据总线、控制总线等。为便于表示，图10中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

[0233] 存储器4003可以是ROM(Read Only Memory，只读存储器)或可存储静态信息和指令的其他类型的静态存储设备，RAM(Random Access Memory，随机存取存储器)或者可存储信息和指令的其他类型的动态存储设备，也可以是EEPROM(Electrically Erasable Programmable Read Only Memory，电可擦可编程只读存储器)、CD-ROM(Compact Disc Read Only Memory，只读光盘)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质、其他磁存储设备、或者能够用于携带或存储计算机程序并能够由计算机读取的任何其他介质，在此不做限定。

[0234] 存储器4003中存储有执行本申请实施例所提供的方法的计算机程序，并可以由处理器4001来控制执行。处理器4001在执行存储器4003中存储的上述计算机程序时，可以实现本申请前述任一方法实施例所示的步骤。

[0235] 本申请实施例还提供了一种计算机可读存储介质，该计算机可读存储介质上存储有计算机程序，计算机程序被处理器执行时可实现本申请前述任一方法实施例的步骤及相应内容。

[0236] 本申请实施例还提供了一种计算机程序产品，该计算机产品中包括计算机程序，计算机程序被处理器执行时可实现本申请前述任一方法实施例的步骤及相应内容。

[0237] 需要说明的是，本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”、“1”、“2”等(如果存在)是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本申请的实施例能够以除图示或文字描述以外的顺序实施。

[0238] 应该理解的是，虽然本申请实施例的流程图中通过箭头指示各个操作步骤，但是这些步骤的实施顺序并不受限于箭头所指示的顺序。除非本文中有明确的说明，否则在本申请实施例的一些实施场景中，各流程图中的实施步骤可以按照需求以其他的顺序执行。此外，各流程图中的部分或全部步骤基于实际的实施场景，可以包括多个子步骤或者多个阶段。这些子步骤或者阶段中的部分或全部可以在同一时刻被执行，这些子步骤或者阶段中的每个子步骤或者阶段也可以分别在不同的时刻被执行。在执行时刻不同的场景下，这些子步骤或者阶段的执行顺序可以根据需求灵活配置，本申请实施例对此不限制。

[0239] 以上所述仅是本申请部分实施场景的可选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请的方案技术构思的前提下，采用基于本申请技术思想的其他类似实施手段，同样属于本申请实施例的保护范畴。

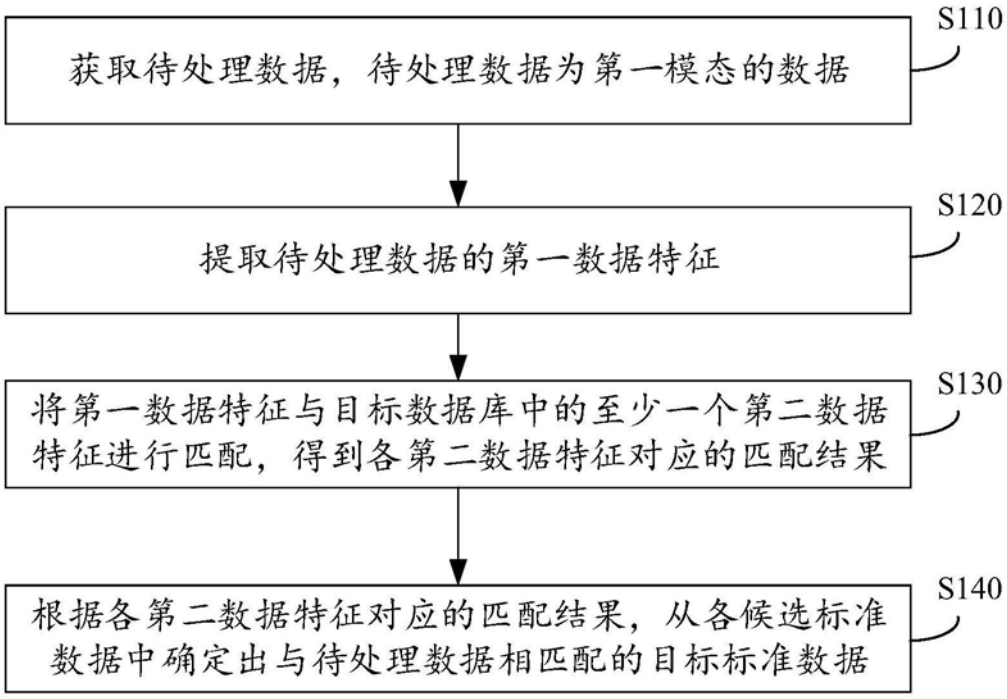

S110
获取待处理数据，待处理数据为第一模态的数据
S120
提取待处理数据的第一数据特征
S130
将第一数据特征与目标数据库中的至少一个第二数据特征进行匹配，得到各第二数据特征对应的匹配结果
S140
根据各第二数据特征对应的匹配结果，从各候选标准数据中确定出与待处理数据相匹配的目标标准数据

图1

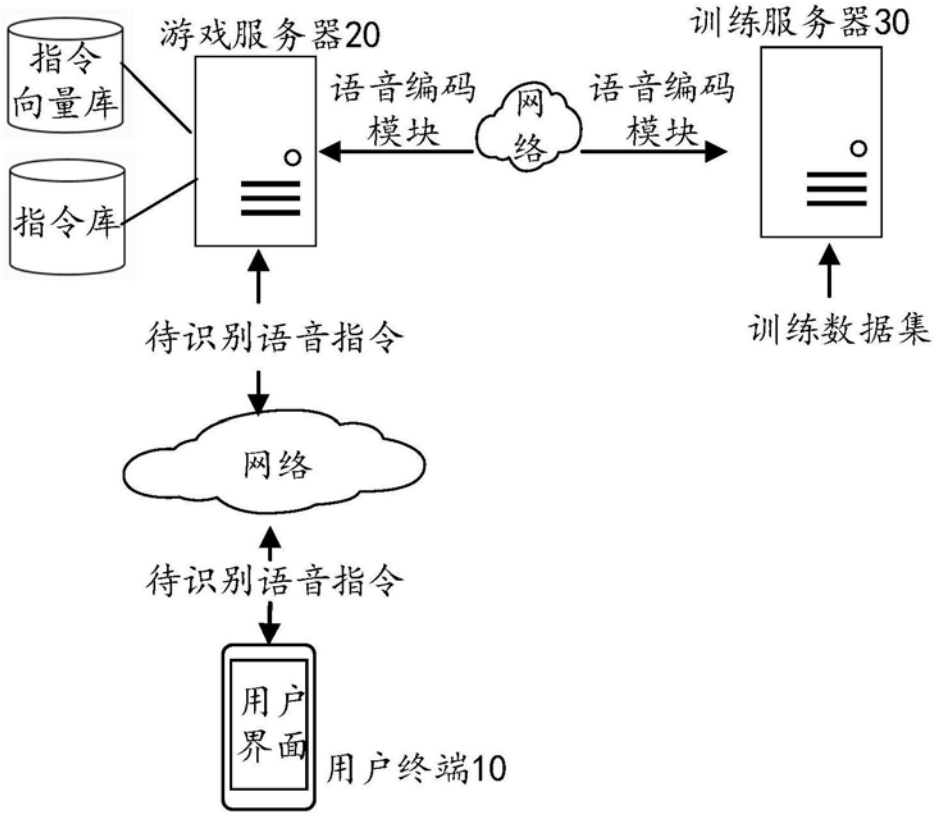

游戏服务器20
训练服务器30
指令向量库
指令库
语音编码模块
网络
语音编码模块
待识别语音指令
训练数据集
网络
待识别语音指令
用户界面
用户终端10

图2

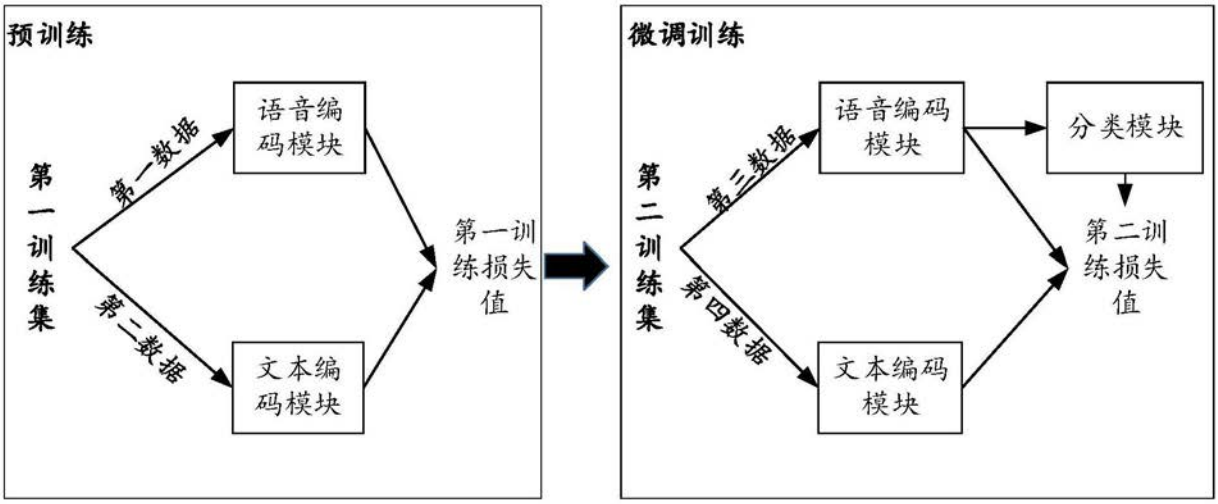

预训练
第一训练集
第一数据
第二数据
语音编码模块
文本编码模块
第一训练损失值
微调训练
第二训练集
第三数据
第四数据
语音编码模块
文本编码模块
分类模块
第二训练损失值

图3

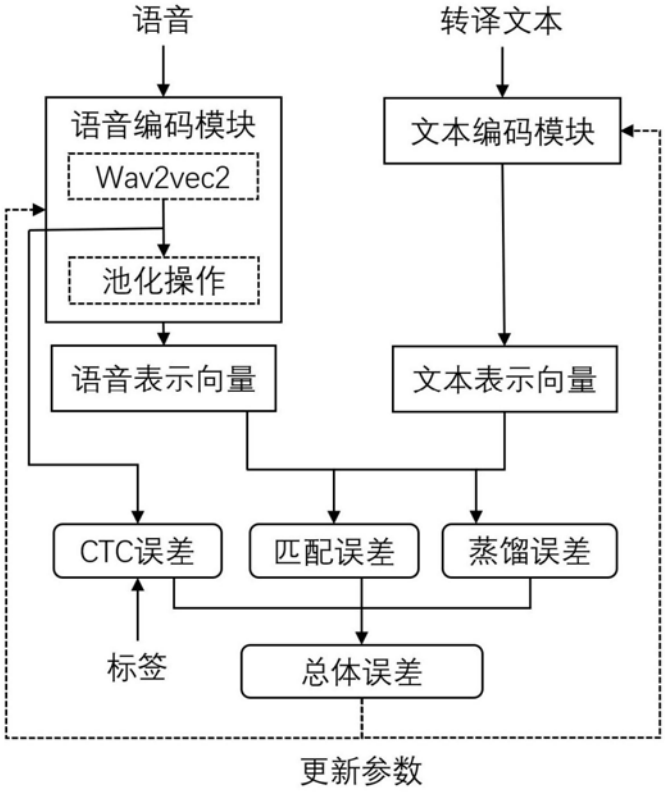

语音
转译文本
语音编码模块
Wav2vec2
池化操作
文本编码模块
语音表示向量
文本表示向量
CTC误差
匹配误差
蒸馏误差
标签
总体误差
更新参数

图4

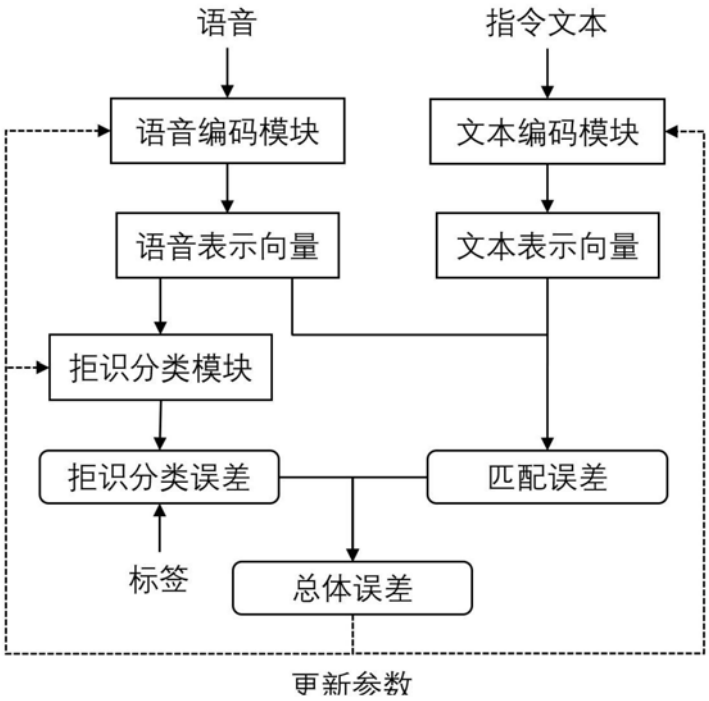

语音
指令文本
语音编码模块
文本编码模块
语音表示向量
文本表示向量
拒识分类模块
拒识分类误差
匹配误差
标签
总体误差
更新参数

图5

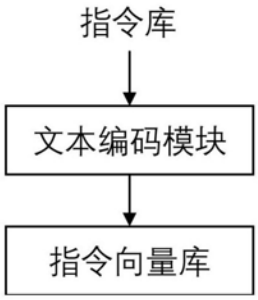

指令库
文本编码模块
指令向量库

图6

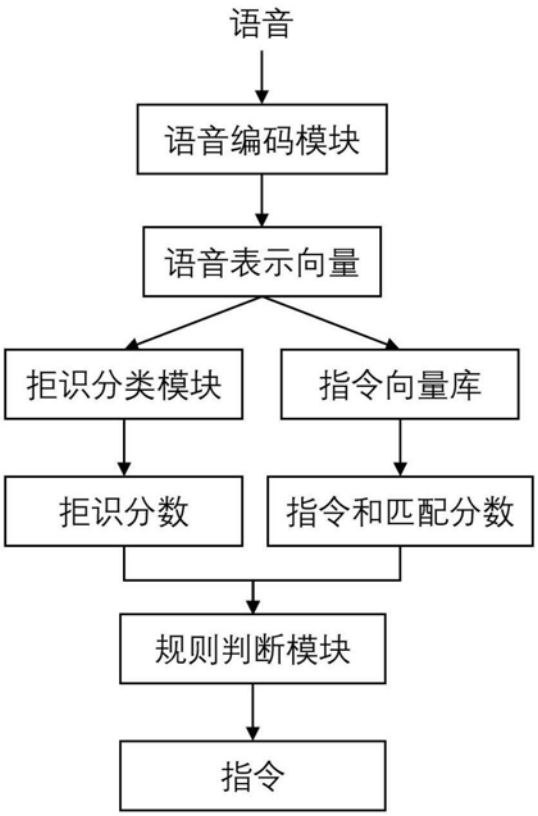

语音
语音编码模块
语音表示向量
拒识分类模块
指令向量库
拒识分数
指令和匹配分数
规则判断模块
指令

图7

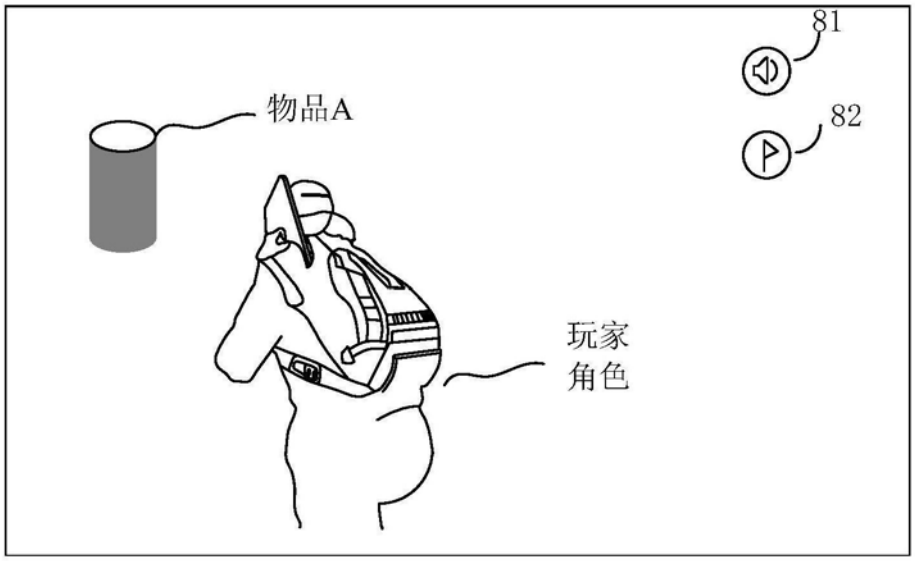

81
82
物品A
玩家
角色

图8a

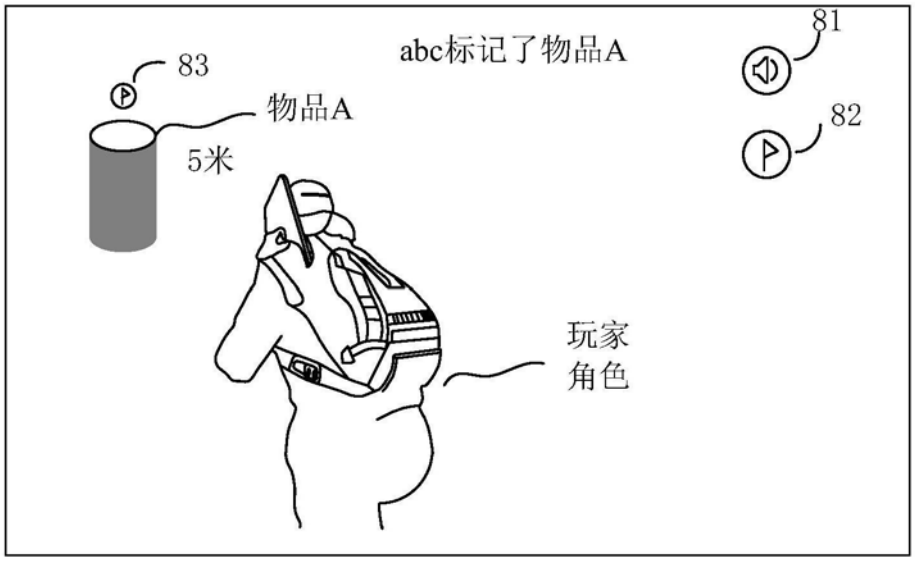

abc标记了物品A
81
83
物品A
5米
82
玩家
角色

图8b

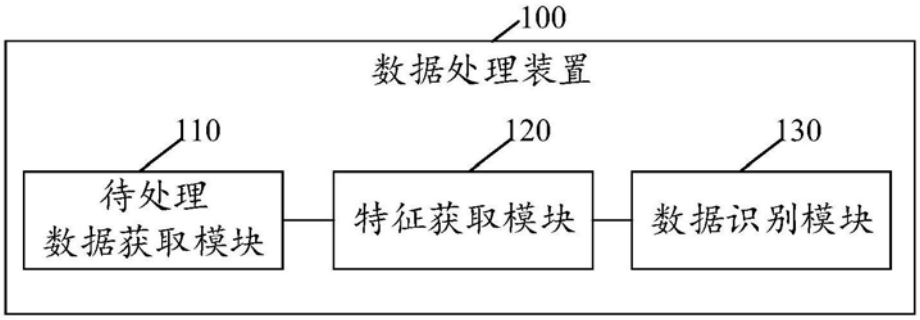

100
数据处理装置
110
120
130
待处理
数据获取模块
特征获取模块
数据识别模块

图9

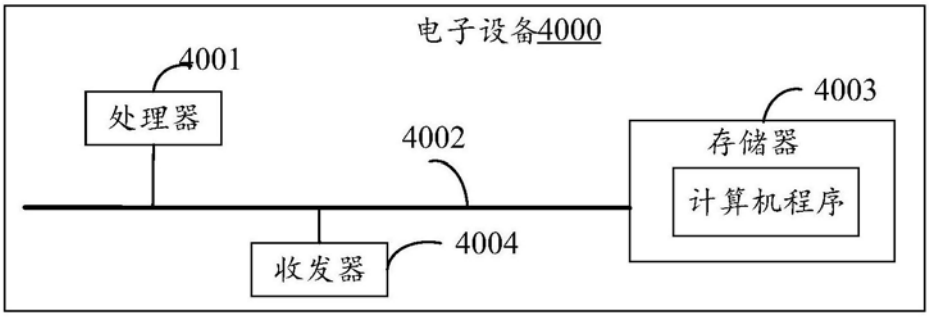

电子设备4000
4001
处理器
4002
4003
存储器
计算机程序
收发器
4004

图10