



(21) 申请号 202310929371.7

G06F 18/214 (2023.01)

(22) 申请日 2023.07.26

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601

(72) 发明人 刘航 冯明超 李浩斌 陈蒙

(74) 专利代理机构 北京律智知识产权代理有限
公司 11438

专利代理师 刘莉

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 40/35 (2020.01)

G06F 40/30 (2020.01)

G06F 18/15 (2023.01)

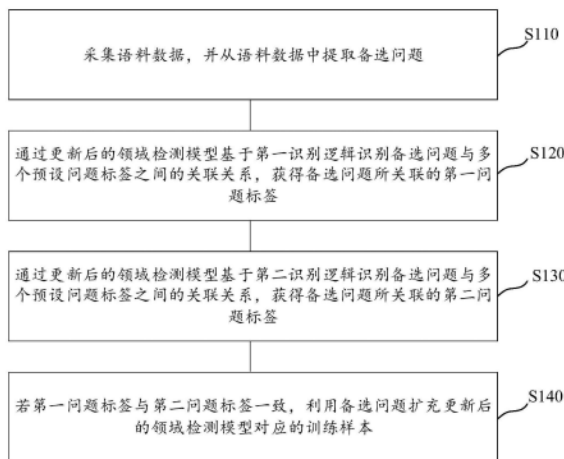
权利要求书4页 说明书19页 附图9页

(54) 发明名称

训练样本扩充方法及装置、计算机存储介
质、电子设备

(57) 摘要

本公开涉及数据处理技术领域,提供了一种训练样本扩充方法、训练样本扩充装置、计算机存储介质、电子设备,其中,训练样本扩充方法包括:采集语料数据,并从语料数据中提取备选问题;通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签;以及,通过更新后的领域检测模型基于第二识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第二问题标签;若第一问题标签与第二问题标签一致,利用备选问题扩充更新后的领域检测模型对应的训练样本。本公开能够提升训练样本的数据质量,从而提升智能问答模型的问答质量。



1. 一种训练样本扩充方法,其特征在于,包括:

采集语料数据,并从所述语料数据中提取备选问题;

通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签;以及,

通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签;

若所述第一问题标签与所述第二问题标签一致,利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

2. 根据权利要求1所述的方法,其特征在于,所述预设问题标签包括所述知识点标签或提问意图标签。

3. 根据权利要求1或2所述的方法,其特征在于,所述通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签,包括:

通过更新后的领域检测模型提取所述备选问题对应的隐特征向量;

获取每个预设问题标签对应的隐特征向量;

计算所述备选问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第一向量距离;

根据多个所述第一向量距离,确定所述备选问题所关联的第一问题标签。

4. 根据权利要求3所述的方法,其特征在于,所述更新后的领域检测模型包括 n 个网络层, n 为大于1的整数;

所述通过更新后的领域检测模型提取所述备选问题对应的隐特征向量,包括:

对所述 n 个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;

计算第 n 个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到 n 个相似度;

将所述 n 个相似度作为所述 n 个网络层对应的输出向量的加权系数,对 n 个所述网络层对应的 n 个输出向量进行加权处理,得到所述备选问题对应的隐特征向量。

5. 根据权利要求3所述的方法,其特征在于,所述根据多个所述第一向量距离,确定所述备选问题所关联的第一问题标签,包括:

从多个所述第一向量距离中筛选出数值最小的第一向量距离,并将所述数值最小的第一向量距离对应的预设问题标签确定为所述备选问题所关联的候选第一问题标签;

查找与所述候选第一问题标签对应的指定距离阈值;

若所述数值最小的第一向量距离大于所述指定距离阈值,确定所述备选问题与所述候选第一问题标签之间不具备关联关系;

若所述数值最小的第一向量距离不大于所述指定距离阈值,将所述候选第一问题标签确定为所述备选问题所关联的第一问题标签。

6. 根据权利要求5所述的方法,其特征在于,在确定所述样本问题与所述候选第一问题标签之间不具备关联关系之后,所述方法还包括:

舍弃所述备选问题。

7. 根据权利要求4所述的方法,其特征在于,所述通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签,包括:

从所述备选问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述备选问题与每个所述预设问题标签的相关性;

根据所述CLS向量,从所述多个预设问题标签中确定所述备选问题所关联的第二问题标签。

8. 根据权利要求1或2所述的方法,其特征在于,所述更新后的领域检测模型通过以下方式训练得到:

获取初始训练样本,所述初始训练样本包括多个预设问题标签,每个所述预设问题标签下关联有多个样本问题;

根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型;

利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本;

将所述清洗后训练样本作为所述训练样本,并根据所述训练样本对所述初始领域检测模型进行迭代训练,得到所述更新后的领域检测模型。

9. 根据权利要求8所述的方法,其特征在于,所述根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型,包括:

将所述初始训练样本输入至所述待训练的领域检测模型中,得到所述待训练的领域检测模型输出的每个所述样本问题所关联的预测问题标签;

根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值;

根据所述损失值,对所述待训练的领域检测模型进行迭代训练,得到所述初始领域检测模型。

10. 根据权利要求9所述的方法,其特征在于,所述根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值,包括:

根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,计算第一类损失值;以及,

根据所述多个样本问题中任意两个样本问题之间的差异程度,计算第二类损失值;

根据所述第一类损失值和所述第二类损失值,确定所述待训练的领域检测模型的损失值。

11. 根据权利要求8所述的方法,其特征在于,所述利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:

利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量;

根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量;

根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签;

根据所述第三问题标签与所述目标样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

12. 根据权利要求11所述的方法,其特征在于,所述初始领域检测模型包括 n 个网络层, n 为大于1的整数;

所述利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量,包括:

对所述 n 个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;

计算第 n 个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到 n 个相似度;

将所述 n 个相似度作为所述 n 个网络层对应的输出向量的加权系数,对 n 个所述网络层对应的 n 个输出向量进行加权处理,得到所述每个样本问题对应的隐特征向量。

13. 根据权利要求11所述的方法,其特征在于,所述根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量,包括:

对所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量进行求均值处理,得到所述每个预设问题标签对应的隐特征向量。

14. 根据权利要求11所述的方法,其特征在于,所述根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签,包括:

从多个所述第二向量距离中筛选出数值最小的第二向量距离,并将所述数值最小的第二向量距离对应的预设问题标签确定为所述样本问题所关联的候选第三问题标签;

计算所述候选第三问题标签对应的指定距离阈值,存储所述候选第三问题标签与所述指定距离阈值之间的对应关系;

若所述数值最小的第二向量距离大于所述指定距离阈值,确定所述样本问题与所述候选第三问题标签之间不具备关联关系;

若所述数值最小的第二向量距离不大于所述指定距离阈值,将所述候选第三问题标签确定为所述样本问题所关联的第三问题标签。

15. 根据权利要求14所述的方法,其特征在于,所述计算所述候选第三问题标签对应的指定距离阈值,包括:

计算所述候选第三问题标签对应的隐特征向量与其他所述预设问题标签对应的隐特征向量之间的多个第三向量距离;

以及,计算所述候选第三问题标签所关联的多个样本问题对应的隐特征向量与所述候选第三问题标签对应的隐特征向量之间的多个第四向量距离;

根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值。

16. 根据权利要求15所述的方法,其特征在于,所述根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值,包括:

从所述多个第三向量距离中筛选数值最小的第三向量距离,从所述多个第四向量距离中筛选出数值最大的第四向量距离;

获取所述数值最小的第三向量距离的目标关联值;

将所述目标关联值与所述数值最大的第四向量距离之间的较大值确定为所述候选第三问题标签对应的指定距离阈值。

17.根据权利要求11所述的方法,其特征在于,所述根据所述第三问题标签与所述样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:

若所述第三问题标签与所述样本问题所关联的预设问题标签不一致,则将所述样本问题从所述初始训练样本中剔除,以得到所述清洗后训练样本;

若所述第三问题标签与所述样本问题所关联的预设问题标签一致,则根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签;

根据所述第三问题标签和所述第四问题标签之间的一致性比对结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

18.根据权利要求17所述的方法,其特征在于,所述根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签,包括:

从所述样本问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述样本问题与每个所述预设问题标签的相关性;

根据所述CLS向量,从所述多个预设问题标签中确定所述样本问题所关联的第四问题标签。

19.一种训练样本扩充装置,其特征在于,包括:

数据采集模块,用于采集语料数据,并从所述语料数据中提取备选问题;

第一识别模块,用于通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签;以及,

第二识别模块,用于通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签;

样本扩充模块,用于若所述第一问题标签与所述第二问题标签一致,利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

20.一种计算机存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1~18中任意一项所述的训练样本扩充方法。

21.一种电子设备,其特征在于,包括:

处理器;以及

存储器,用于存储所述处理器的可执行指令;

其中,所述处理器配置为经由执行所述可执行指令来执行权利要求1~18中任意一项所述的训练样本扩充方法。

训练样本扩充方法及装置、计算机存储介质、电子设备

技术领域

[0001] 本公开涉及数据处理技术领域，特别涉及一种训练样本扩充方法、训练样本扩充装置、计算机存储介质及电子设备。

背景技术

[0002] 智能问答模型以一问一答形式，精确的定位网站用户所需要的提问知识，通过与网站用户进行交互，为网站用户提供个性化的信息服务，而智能问答模型的问答效果依赖于训练数据的质量，一般而言，训练数据的质量越高，错误数据越少，智能问答模型的问答效果就越好。

[0003] 目前，一般是通过人工收集知识点以及每个知识点下的问题，构建并扩充智能问答模型的训练样本。然而，上述方案会导致训练样本的数据质量越来越低，从而影响智能问答模型的问答效果。

[0004] 鉴于此，本领域亟需开发一种新的训练样本扩充方法及装置。

[0005] 需要说明的是，上述背景技术部分公开的信息仅用于加强对本公开的背景的理解。

发明内容

[0006] 本公开的目的在于提供一种训练样本扩充方法、训练样本扩充装置、计算机存储介质及电子设备，进而至少在一定程度上克服相关技术中由于训练样本质量低下而导致的影响智能问答模型的问答效果的技术问题。

[0007] 本公开的其他特性和优点将通过下面的详细描述变得显然，或部分地通过本公开的实践而习得。

[0008] 根据本公开的第一方面，提供一种训练样本扩充方法，包括：采集语料数据，并从所述语料数据中提取备选问题；通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系，获得所述备选问题所关联的第一问题标签；以及，通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系，获得所述备选问题所关联的第二问题标签；若所述第一问题标签与所述第二问题标签一致，利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

[0009] 在本公开的示例性实施例中，所述预设问题标签包括所述知识点标签或提问意图标签。

[0010] 在本公开的示例性实施例中，所述通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系，获得所述备选问题所关联的第一问题标签，包括：通过更新后的领域检测模型提取所述备选问题对应的隐特征向量；获取每个预设问题标签对应的隐特征向量；计算所述备选问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第一向量距离；根据多个所述第一向量距离，确定所述

备选问题所关联的第一问题标签。

[0011] 在本公开的示例性实施例中,所述更新后的领域检测模型包括 n 个网络层, n 为大于1的整数;所述通过更新后的领域检测模型提取所述备选问题对应的隐特征向量,包括:对所述 n 个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;计算第 n 个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到 n 个相似度;将所述 n 个相似度作为所述 n 个网络层对应的输出向量的加权系数,对 n 个所述网络层对应的 n 个输出向量进行加权处理,得到所述备选问题对应的隐特征向量。

[0012] 在本公开的示例性实施例中,所述根据多个所述第一向量距离,确定所述备选问题所关联的第一问题标签,包括:从多个所述第一向量距离中筛选出数值最小的第一向量距离,并将所述数值最小的第一向量距离对应的预设问题标签确定为所述备选问题所关联的候选第一问题标签;查找与所述候选第一问题标签对应的指定距离阈值;若所述数值最小的第一向量距离大于所述指定距离阈值,确定所述备选问题与所述候选第一问题标签之间不具备关联关系;若所述数值最小的第一向量距离不大于所述指定距离阈值,将所述候选第一问题标签确定为所述备选问题所关联的第一问题标签。

[0013] 在本公开的示例性实施例中,在确定所述样本问题与所述候选第一问题标签之间不具备关联关系之后,所述方法还包括:舍弃所述备选问题。

[0014] 在本公开的示例性实施例中,所述通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签,包括:从所述备选问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述备选问题与每个所述预设问题标签的相关性;根据所述CLS向量,从所述多个预设问题标签中确定所述备选问题所关联的第二问题标签。

[0015] 在本公开的示例性实施例中,所述更新后的领域检测模型通过以下方式训练得到:获取初始训练样本,所述初始训练样本包括多个预设问题标签,每个所述预设问题标签下关联有多个样本问题;根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型;利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本;将所述清洗后训练样本作为所述训练样本,并根据所述训练样本对所述初始领域检测模型进行迭代训练,得到所述更新后的领域检测模型。

[0016] 在本公开的示例性实施例中,所述根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型,包括:将所述初始训练样本输入至所述待训练的领域检测模型中,得到所述待训练的领域检测模型输出的每个所述样本问题所关联的预测问题标签;根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值;根据所述损失值,对所述待训练的领域检测模型进行迭代训练,得到所述初始领域检测模型。

[0017] 在本公开的示例性实施例中,所述根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值,包括:根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,计算第一类损失值;以及,根据所述多个样本问题中任意两个样本问题之间的差异程度,计算第二类损失值;根据所述第一类损失值

和所述第二类损失值,确定所述待训练的领域检测模型的损失值。

[0018] 在本公开的示例性实施例中,所述利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量;根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量;根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签;根据所述第三问题标签与所述目标样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

[0019] 在本公开的示例性实施例中,所述初始领域检测模型包括n个网络层,n为大于1的整数;所述利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量,包括:对所述n个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;计算第n个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到n个相似度;将所述n个相似度作为所述n个网络层对应的输出向量的加权系数,对n个所述网络层对应的n个输出向量进行加权处理,得到所述每个样本问题对应的隐特征向量。

[0020] 在本公开的示例性实施例中,所述根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量,包括:对所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量进行求均值处理,得到所述每个预设问题标签对应的隐特征向量。

[0021] 在本公开的示例性实施例中,所述根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签,包括:从多个所述第二向量距离中筛选出数值最小的第二向量距离,并将所述数值最小的第二向量距离对应的预设问题标签确定为所述样本问题所关联的候选第三问题标签;计算所述候选第三问题标签对应的指定距离阈值,存储所述候选第三问题标签与所述指定距离阈值之间的对应关系;若所述数值最小的第二向量距离大于所述指定距离阈值,确定所述样本问题与所述候选第三问题标签之间不具备关联关系;若所述数值最小的第二向量距离不大于所述指定距离阈值,将所述候选第三问题标签确定为所述样本问题所关联的第三问题标签。

[0022] 在本公开的示例性实施例中,所述计算所述候选第三问题标签对应的指定距离阈值,包括:计算所述候选第三问题标签对应的隐特征向量与其他所述预设问题标签对应的隐特征向量之间的多个第三向量距离;以及,计算所述候选第三问题标签所关联的多个样本问题对应的隐特征向量与所述候选第三问题标签对应的隐特征向量之间的多个第四向量距离;根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值。

[0023] 在本公开的示例性实施例中,所述根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值,包括:从所述多个第三向量距离中筛选数值最小的第三向量距离,从所述多个第四向量距离中筛选出数值最大的第四向量距离;获取所述数值最小的第三向量距离的目标关联值;将所述目标关联值与所述数值最大的第四向量距离之间的较大值确定为所述候选第三问题标签对应的指定距离阈值。

[0024] 在本公开的示例性实施例中,所述根据所述第三问题标签与所述样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:若所述第三问题标签与所述样本问题所关联的预设问题标签不一致,则将所述样本问题从所述初始训练样本中剔除,以得到所述清洗后训练样本;若所述第三问题标签与所述样本问题所关联的预设问题标签一致,则根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签;根据所述第三问题标签和所述第四问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

[0025] 在本公开的示例性实施例中,所述根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签,包括:从所述样本问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述样本问题与每个所述预设问题标签的相关性;根据所述CLS向量,从所述多个预设问题标签中确定所述样本问题所关联的第四问题标签。

[0026] 根据本公开的第二方面,提供一种训练样本扩充装置,包括:数据采集模块,用于采集语料数据,并从所述语料数据中提取备选问题;第一识别模块,用于通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签;以及,第二识别模块,用于通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签;样本扩充模块,用于若所述第一问题标签与所述第二问题标签一致,利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

[0027] 根据本公开的第三方面,提供一种计算机存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述第一方面所述的训练样本扩充方法。

[0028] 根据本公开的第四方面,提供一种电子设备,包括:处理器;以及存储器,用于存储所述处理器的可执行指令;其中,所述处理器配置为经由执行所述可执行指令来执行上述第一方面所述的训练样本扩充方法。

[0029] 由上述技术方案可知,本公开示例性实施例中的训练样本扩充方法、训练样本扩充装置、计算机存储介质及电子设备至少具备以下优点和积极效果:

[0030] 在本公开的一些实施例所提供的技术方案中,一方面,通过采集语料数据,并从语料数据中提取备选问题,提供了一种新的训练样本的扩充思路,使得除了对训练样本本身进行处理实现数据扩充之外,还可以基于采集到的语料数据对训练样本进行扩充。进一步的,通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签,以及,通过更新后的领域检测模型基于第二识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第二问题标签,能够通过两种识别逻辑准确确定出备选问题与多个预设问题标签之间的关联关系。另一方面,若第一问题标签与第二问题标签一致,利用备选问题扩充更新后的领域检测模型对应的训练样本,能够在准确理清备选问题与多个预设问题标签之间的关联关系之后,再基于备选问题扩充训练样本,从而,解决了相关技术中在样本扩充时人工划分关联关系所导致的出现混淆或错误数据的问题,提升扩充之后的训练样本的数据质量,从而,进一步提升智能问答系统的问答效果。

[0031] 本公开应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0032] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0033] 图1示出本公开实施例中训练样本扩充方法的流程示意图;

[0034] 图2示出本公开实施例中如何训练得到更新后的领域检测模型的流程示意图;

[0035] 图3示出本公开实施例中如何根据初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型的流程示意图;

[0036] 图4示出本公开实施例中如何根据预测问题标签与样本问题所关联的预设问题标签之间的差异程度,确定待训练的领域检测模型的损失值的流程示意图;

[0037] 图5示出本公开实施例中如何利用初始领域检测模型对初始训练样本进行数据清洗,得到清洗后训练样本的流程示意图;

[0038] 图6示出本公开实施例中如何利用初始领域检测模型获取每个样本问题对应的隐特征向量的流程示意图;

[0039] 图7示出本公开实施例中如何利用初始领域检测模型获取每个样本问题对应的隐特征向量的示意图;

[0040] 图8示出本公开实施例中如何根据每个样本问题对应的隐特征向量与每个预设问题标签对应的隐特征向量之间的第二向量距离,确定每个样本问题所关联的第三问题标签的流程示意图;

[0041] 图9示出本公开实施例中如何确定样本问题所关联的候选第三问题标签的示意图;

[0042] 图10示出本公开实施例中如何确定预设问题标签A对应的指定距离阈值的示意图;

[0043] 图11示出本公开实施例中如何根据第三问题标签与样本问题所关联的预设问题标签之间的一致性对比结果,对初始训练样本进行数据清洗,得到清洗后训练样本的流程示意图;

[0044] 图12示出本公开实施例中如何通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签的流程示意图;

[0045] 图13示出本公开实施例中如何根据多个第一向量距离,确定备选问题所关联的第一问题标签的流程示意图;

[0046] 图14示出本公开实施例中训练样本扩充方法的整体示意图;

[0047] 图15示出本公开示例性实施例中训练样本扩充装置的结构示意图;

[0048] 图16示出本公开示例性实施例中电子设备的结构示意图。

具体实施方式

[0049] 现在将参考附图更全面地描述示例实施方式。然而,示例实施方式能够以多种形式实施,且不应被理解为限于在此阐述的范例;相反,提供这些实施方式使得本公开将更加

全面和完整,并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中,提供许多具体细节从而给出对本公开的实施方式的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而省略所述特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知技术方案以避免喧宾夺主而使得本公开的各方面变得模糊。

[0050] 本说明书中使用用语“一个”、“一”、“该”和“所述”用以表示存在一个或多个要素/组成部分/等;用语“包括”和“具有”用以表示开放式的包括在内的意思并且是指除了列出的要素/组成部分/等之外还可存在另外的要素/组成部分/等;用语“第一”和“第二”等仅作为标记使用,不是对其对象的数量限制。

[0051] 此外,附图仅为本公开的示意性图解,并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体,不一定必须与物理或逻辑上独立的实体相对应。

[0052] 智能问答系统包含知识库应答和意图应答两个核心模块,各个核心模块中包含的问题数量和问题之间的混淆度,直接决定了该核心模块的问答效果。

[0053] 其中,上述知识库由多个知识点组成,每个知识点下有一个对应的答案以及多个问题。在知识库应答时,用户问题会与知识库的问题进行匹配,选择出匹配度最高的问题,返回该问题所对应知识点的答案。在匹配系统保持固定的情况下,知识点下的问题丰富度越高,混淆度越低(与其他知识点没有问题语义上的重合),应答的效果越好。但是,因为每个知识点下的问题通常都是人工收集的,而随着知识点数量的增多,人工区分知识点含义的界线变得越来越模糊,知识点下的问题也逐渐变得容易混淆。

[0054] 意图应答采用了一批意图语料作为训练数据(训练数据通常由多个意图组成,每个意图下有多个训练语料),离线训练一个分类模型,然后将模型部署至系统中。当用户问题进入到意图模块时,该模型会对用户问题进行分类识别,判断出用户提问的意图,然后进行应答。在这个过程中,意图模型发挥着重要作用,其准确性直接影响着应答效果。而模型的准确性依赖于训练数据的质量,当训练数据质量越高,错误数据越少,模型的效果也就越好。因此,只有在保证这些语料所对应的意图是正确的情况下,训练出的模型才能具有较好的意图识别效果。然而,目前的训练数据一般都是人工标注得到的,而人工标注成本较高,并且,人工标注数据可能会出现标注错误的情况。

[0055] 目前,常用的数据增强方案都是基于已有的数据进行扩充的,例如:基于生成式方法生成一个与已有问题语义一致的新问题,或者,基于替换式方法将已有问法中的词汇进行同义词替换,生成一个语义相同的新问题。

[0056] 上述方案虽然可以丰富问题的表达形式,但无法保证问题的流畅度,可能会出现不符合口语化表达的文本。同时,已有数据中的混淆问题和错误标注问题依然无法得到有效解决,反而在数据扩充后会进一步增加混淆样本和错误样本的数量,导致训练样本的数据质量较低,智能问答模型的问答效果无法取得明显改善。

[0057] 在本公开的实施例中,首先提供了一种训练样本扩充方法,至少在一定程度上克服相关技术中由于训练样本质量低下而导致的影响智能问答模型的问答效果的技术问题的缺陷。

[0058] 图1示出本公开实施例中训练样本扩充方法的流程示意图,该训练样本扩充方法的执行主体可以是对训练样本进行扩充处理的服务器。

[0059] 参考图1,根据本公开的一个实施例的训练样本扩充方法包括以下步骤:

[0060] 步骤S110,采集语料数据,并从语料数据中提取备选问题;

[0061] 步骤S120,通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签;以及,

[0062] 步骤S130,通过更新后的领域检测模型基于第二识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第二问题标签;

[0063] 步骤S140,若第一问题标签与第二问题标签一致,利用备选问题扩充更新后的领域检测模型对应的训练样本。

[0064] 在图1所示实施例所提供的技术方案中,一方面,通过采集语料数据,并从语料数据中提取备选问题,提供了一种新的训练样本的扩充思路,使得除了对训练样本本身进行处理实现数据扩充之外,还可以基于采集到的语料数据对训练样本进行扩充。进一步的,通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签,以及,通过更新后的领域检测模型基于第二识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第二问题标签,能够通过两种识别逻辑准确确定出备选问题与多个预设问题标签之间的关联关系。另一方面,若第一问题标签与第二问题标签一致,利用备选问题扩充更新后的领域检测模型对应的训练样本,能够在准确理清备选问题与多个预设问题标签之间的关联关系之后,再基于备选问题扩充训练样本,从而,解决了相关技术中在样本扩充时人工划分关联关系所导致的出现混淆或错误数据的问题,提升扩充之后的训练样本的数据质量,从而,进一步提升智能问答系统的问答效果。

[0065] 以下对图1中的各个步骤的具体实现过程进行详细阐述:

[0066] 在步骤S110之前,需要说明的是,本公开可以先训练一更新后的领域检测模型,该更新后的领域检测模型的作用在于:通过两种识别逻辑识别出问题所关联的第一问题标签与第二问题标签。

[0067] 参考图2,图2示出本公开实施例中如何训练得到更新后的领域检测模型的流程示意图,包含步骤S201-步骤S204:

[0068] 在步骤S201中,获取初始训练样本。

[0069] 本步骤中,初始训练样本可以是目前现存的用于智能问答的知识库中包含的数据,或者,可以是目前现存的用于意图应答的训练语料。

[0070] 上述初始训练样本包括多个预设问题标签,每个预设问题标签下关联有多个样本问题。

[0071] 示例性的,针对知识库应答的形式,初始训练样本可以包含多个知识点标签(即预设问题标签),每个知识点标签下关联多个样本问题,例如:训练样本包含知识点标签A、知识点标签B和知识点标签C,而知识点标签A下关联样本问题A1、样本问题A2和样本问题A3,知识点标签B下关联样本问题B1、样本问题B2、样本问题B3和样本问题B4,知识点标签C下关联样本问题C1和样本问题C2。

[0072] 而针对意图应答的形式,初始训练样本可以包含多个提问意图标签(即预设问题

标签),每个提问意图标签下关联多个样本问题,例如:训练样本包含提问意图标签A、提问意图标签B和提问意图标签C,而提问意图标签A下关联样本问题A1、样本问题A2和样本问题A3,提问意图标签B下关联样本问题B1、样本问题B2、样本问题B3和样本问题B4,提问意图标签C下关联样本问题C1和样本问题C2。

[0073] 在获取到初始训练样本之后,可以进入步骤S202中,根据初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型。

[0074] 本步骤中,参考图3,图3示出本公开实施例中如何根据初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型的流程示意图,包含步骤S301-步骤S303:

[0075] 在步骤S301中,将初始训练样本输入至待训练的领域检测模型中,得到待训练的领域检测模型输出的每个样本问题所关联的预测问题标签。

[0076] 本步骤中,可以将上述初始训练样本输入至上述待训练的领域检测模型中,根据上述待训练的领域检测模型的输出,得到每个样本问题所关联的预测问题标签。其中,上述待训练的领域检测模型可以是BERT (Bidirectional Encoder Representations from Transformers,来自变换器的双向编码器表征量)模型,可以根据实际情况自行设定,本公开对此不作特殊限定。

[0077] 示例性的,样本问题所关联的预测问题标签可以是样本问题所属的知识点标签,或者,样本问题所对应的提问意图标签。

[0078] 在步骤S302中,根据预测问题标签与样本问题所关联的预设问题标签之间的差异程度,以及,多个样本问题中任意两个样本问题之间的差异程度,确定待训练的领域检测模型的损失值。

[0079] 本步骤中,参考图4,图4示出本公开实施例中如何根据预测问题标签与样本问题所关联的预设问题标签之间的差异程度,确定待训练的领域检测模型的损失值的流程示意图,包含步骤S401-步骤S403:

[0080] 在步骤S401中,根据预测问题标签与样本问题所关联的预设问题标签之间的差异程度,计算第一类损失值。

[0081] 本步骤中,第一类损失值可以是交叉熵损失值 (Cross Entropy,CE),交叉熵损失是表征真实样本标签与预测样本标签之间的差值。

[0082] 在步骤S402中,根据多个样本问题中任意两个样本问题之间的差异程度,计算第二类损失值。

[0083] 本步骤中,第二类损失值可以是对比损失 (Contrastive Loss,CL),对比损失是一种二元分类损失函数,通常用于度量两个样本之间的相似度,根据相似度判断两个样本是否属于同一类别,若属于同一类别,则对比损失为0,如果属于不同的类别,则对比损失为一个正数。

[0084] 在步骤S403中,根据第一类损失值和第二类损失值,确定待训练的领域检测模型的损失值。

[0085] 本步骤中,可以将上述交叉熵损失值 and 对比损失值进行累加,并将得到的累加值作为待训练的领域检测模型的损失值。

[0086] 接着参考图3,在步骤S303中,根据损失值,对待训练的领域检测模型进行迭代训

练,得到初始领域检测模型。

[0087] 本步骤中,可以根据上述初始训练样本对待训练的领域检测模型进行迭代训练,直至上述待训练的领域检测模型的损失值趋于收敛,得到初始领域检测模型。

[0088] 在得到初始领域检测模型之后,鉴于上述初始领域检测模型是根据初始训练样本训练得到的,而初始训练样本中的数据是人工标记得到的,因而,可能会存在一些错误或问题与所属问题标签搞混的混淆数据,影响智能问答模型的问答质量,从而,可以利用上述初始检测模型对上述初始训练样本进行数据清洗,以提升样本质量。

[0089] 接着参考图2,在步骤S203中,利用初始领域检测模型对初始训练样本进行数据清洗,得到清洗后训练样本。

[0090] 本步骤中,参考图5,图5示出本公开实施例中如何利用初始领域检测模型对初始训练样本进行数据清洗,得到清洗后训练样本的流程示意图,包含步骤S501-步骤S503:

[0091] 在步骤S501中,利用初始领域检测模型获取每个样本问题对应的隐特征向量。

[0092] 本步骤中,上述初始领域检测模型可以包含多个网络层,从而,参考图6,图6示出本公开实施例中如何利用初始领域检测模型获取每个样本问题对应的隐特征向量的流程示意图,包含步骤S601-步骤S603:

[0093] 在步骤S601中,对多个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量。

[0094] 本步骤中,上述初始检测模型的每个网络层可以对上述样本问题进行编码,从而,每个网络层对应各自的输出特征,之后,可以对每个网络层的输出特征进行池化处理(mean-pooling),得到每个网络层对应的输出向量 h_j ($j=1,2,3\cdots n$)。

[0095] 在步骤S602中,计算第多个网络层对应的输出向量与每个网络层对应的输出向量之间的相似度,得到多个相似度。

[0096] 本步骤中,上述相似度可以是余弦相似度,从而,在得到每个网络层对应的输出向量之后,可以基于以下公式1计算第多个网络层对应的输出向量 h_n 与每个网络层对应的输出向量 h_j 之间的余弦相似度 s_j :

$$[0097] \quad s_j = \frac{h_n \cdot h_j}{|h_n| |h_j|} \quad \text{公式 1}$$

[0098] 需要说明的是,在得到上述余弦相似度 s_j 之后,还可以对基于以下公式2对上述余弦相似度 s_j 进行归一化处理:

$$[0099] \quad \alpha_j = \frac{e^{s_j}}{\sum_{m=1}^n e^{s_m}} \quad \text{公式 2}$$

[0100] 其中, α_j 表示归一化处理之后的余弦相似度, e 表示自然对数。

[0101] 在步骤S603中,将多个相似度作为多个网络层对应的输出向量的加权系数,对多个网络层对应的多个输出向量进行加权处理,得到每个样本问题对应的隐特征向量。

[0102] 本步骤中,在得到上述多个相似度之后,可以将上述多个相似度作为多个网络层对应的输出向量的加权系数,进而,根据上述多个加权系数对上述多个输出向量进行加权处理,得到每个样本问题对应的隐特征向量 $\bar{X}_i$ (i 为从1至上述多个样本问题的总数目的正整数)。具体的,参考以下公式3:

$$[0103] \quad \bar{X}_i = \sum_{j=1}^n \alpha_j \cdot h_j \quad \text{公式 3}$$

[0104] 参考图7,图7示出本公开实施例中如何利用初始领域检测模型获取每个样本问题对应的隐特征向量的示意图,如图7所示:

[0105] 可以先对每个预设问题标签下包含的多个样本问题(举例而言,“退款”标签下包含的样本问题可以是:我不想要了咋拒收退款;拒收怎么退款;“退货”标签下包含的样本问题可以是:退货地址给我;有退货地址吗)输入至上述初始领域检测模型中,从而,初始领域检测模型的多个网络层(Layer₁至Layer_n)可以输出各自对应的输出特征,针对每个网络层对应的输出特征,可以进行池化处理(mean-pooling),得到多个样本问题对应的隐特征向量 $\bar{X}_1$ 、 $\bar{X}_2$ …… $\bar{X}_{i-1}$, $\bar{X}_i$ 。

[0106] 在得到训练样本中的多个样本问题对应的隐特征向量之后,可以进入步骤S502中,根据每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定每个预设问题标签对应的隐特征向量。

[0107] 本步骤中,可以确定每个预设问题标签下所关联的目标样本问题对应的隐特征向量,进而,对该多个样本问题对应的隐特征向量进行求均值处理,得到每个预设问题标签对应的隐特征向量。示例性的,参考上述步骤S201的相关解释,以预设问题标签为知识点标签A为例进行说明,则在确定上述知识点标签A对应的隐特征向量时,可以对上述样本问题A1对应的隐特征向量、上述样本问题A2对应的隐特征向量、上述样本问题A3对应的隐特征向量每个维度的值进行求均值处理,得到上述知识点标签A对应的隐特征向量。

[0108] 在计算出每个预设问题标签对应的隐特征向量之后,可以将其进行对应存储,以便后续模型应用过程中快速调用该数据。

[0109] 在步骤S503中,根据样本问题对应的隐特征向量与每个预设问题标签对应的隐特征向量之间的第二向量距离,确定每个样本问题所关联的第三问题标签。

[0110] 本步骤中,上述第二向量距离可以是Mahalanobis距离(马氏距离,从数学上来看,马氏距离是表示数据的协方差距离,它是一种有效的计算两个未知样本集的相似度的方法)、欧式距离、曼哈顿距离、切比雪夫距离等,可以根据实际情况自行设定,本公开对此不作特殊限定。

[0111] 参考图8,图8示出本公开实施例中如何根据每个样本问题对应的隐特征向量与每个预设问题标签对应的隐特征向量之间的第二向量距离,确定每个样本问题所关联的第三问题标签的流程示意图,包含步骤S801-步骤S804:

[0112] 在步骤S801中,从多个第二向量距离中筛选出数值最小的第二向量距离,并将数值最小的第二向量距离对应的预设问题标签确定为样本问题所关联的候选第三问题标签。

[0113] 本步骤中,在计算得到多个第二向量距离之后,可以将上述多个第二向量距离进行数值比对,以从上述多个第二向量距离中筛选出数值最小的第二向量距离,进而,可以将数值最小的第二向量距离对应的预设问题标签确定为样本问题所关联的候选第三问题标签。

[0114] 示例性的,参考图9,图9示出本公开实施例中如何确定样本问题所关联的候选第三问题标签的示意图,如图9所示:

[0115] 假设计算出来样本问题对应的隐特征向量X与预设问题标签A之间的第二向量距

离为 l_1 ,样本问题对应的隐特征向量X与预设问题标签B之间的第二向量距离为 l_2 ,样本问题对应的隐特征向量X与预设问题标签C之间的第二向量距离为 l_3 ,且 $l_1 < l_2 < l_3$,则可以将数值最小的第二向量距离 l_1 对应的预设问题标签A确定为上述样本问题对应的候选第三问题标签。

[0116] 接着参考图8,在步骤S802中,计算候选第三问题标签对应的指定距离阈值,存储候选第三问题标签与指定距离阈值之间的对应关系。

[0117] 本步骤中,首先,可以先计算候选第三问题标签对应的隐特征向量与其他预设问题标签对应的隐特征向量之间的多个第三向量距离(可以是马氏距离),该第三向量距离用于衡量各个候选第三问题标签与其他预设问题标签之间的相关性,即类间距离。

[0118] 其次,可以计算候选第三问题标签下所关联的多个样本问题对应的隐特征向量与候选第三问题标签对应的隐特征向量之间的多个第四向量距离(也可以是马氏距离),该第四向量距离用于衡量候选第三问题标签与其所关联的多个样本问题之间的相关性,即类内距离。

[0119] 再次,可以从上述多个第三向量距离中筛选出数值最小的第三向量距离,即类间距离的最小值,并获取其目标关联值,示例性的,目标关联值可以是上述数值最小的第三向量距离的中值(举例而言,假设数值最小的第二向量距离为 r ,那么,目标关联值可以取 $\frac{r}{2}$)。

以及,从上述多个第四向量距离中筛选出数值最大的第四向量距离,即类内距离的最大值。之后,可以将上述目标关联值和上述数值最大的第四向量距离之间的较大值确定为候选第三问题标签对应的指定距离阈值。

[0120] 示例性的,参考图10,图10示出本公开实施例中如何确定预设问题标签A对应的指定距离阈值的示意图,如图10所示:

[0121] 参考上述图9的相关解释,在将预设问题标签A确定为上述样本问题对应的候选第三问题标签之后,此时可以计算预设问题标签A的隐特征向量与其他预设问题标签的隐特征向量之间的多个第三向量距离,从多个第三向量距离中筛选出数值最小的第三向量距离,例如:图10中所示的 l_4 ,进而,计算候选第三问题标签的隐特征向量与其包含的多个样本问题的隐特征向量之间的多个第四向量距离,从多个第四向量距离中筛选出数值最大的第四向量距离,例如,图10中所示的 l_5 ,进而,将 $\frac{l_4}{2}$ 确定上述目标关联值,将 $\frac{l_4}{2}$ 和 l_5 之间的较大值确定为上述指定距离阈值。

[0122] 需要说明的是,可以在遍历上述样本问题的时候,计算出每个预设问题标签对应的指定距离阈值,并将其存储备用,便于后续快速查找每个预设问题标签对应的指定距离阈值。

[0123] 在步骤S803中,若数值最小的第二向量距离大于指定距离阈值,确定样本问题与候选第三问题标签之间不具备关联关系。

[0124] 本步骤中,可以比较上述数值最小的第二向量距离与上述指定距离阈值之间的数值关系,若上述数值最小的第二向量距离大于上述指定距离阈值,则可以确定样本问题与上述候选第三问题标签之间不具备关联关系,此时,可以将上述样本问题从上述训练样本中剔除。

[0125] 在步骤S804中,若数值最小的第二向量距离不大于指定距离阈值,将候选第三问

题标签确定为样本问题所关联的第三问题标签。

[0126] 本步骤中,若上述数值最小的第二向量距离不大于上述指定距离阈值,则可以将上述候选第三问题标签确定为上述样本问题所关联的第三问题标签。

[0127] 在确定出样本问题所关联的第三问题标签之后,可以接着参考图5,在步骤S504中,根据第三问题标签与样本问题所关联的预设问题标签之间的一致性对比结果,对初始训练样本进行数据清洗,得到清洗后训练样本。

[0128] 本步骤中,可以参考图11,图11示出本公开实施例中如何根据第三问题标签与样本问题所关联的预设问题标签之间的一致性对比结果,对初始训练样本进行数据清洗,得到清洗后训练样本的流程示意图,包含步骤S1101-步骤S1103:

[0129] 在步骤S1101中,若第三问题标签与样本问题所关联的预设问题标签不一致,则将样本问题从初始训练样本中剔除,以得到清洗后训练样本。

[0130] 本步骤中,可以将上述第三问题标签与上述样本问题所关联的预设问题标签进行比对,若二者不一致,则可以直接将上述样本问题从上述初始训练样本中剔除,以得到清洗后训练样本。示例性的,假设上述步骤S804中确定出来样本问题A1所对应的第三问题标签为知识点标签B,而上述样本问题A1所对应的预设问题标签为知识点标签A时,则可以确定出二者不一致,此时,说明样本问题A1存在归类错误的可能性,从而,可以将样本问题A1从初始训练样本中剔除。

[0131] 在步骤S1102中,若第三问题标签与样本问题所关联的预设问题标签一致,则根据每个样本问题对应的隐特征向量,识别样本问题所关联的第四问题标签。

[0132] 本步骤中,如果上述第三问题标签与上述样本问题所关联的预设问题标签一致,则可以从样本问题对应的隐特征向量中抽取CLS向量,CLS向量用于表征样本问题与每个预设问题标签的相关性,示例性的,上述CLS向量可以表征样本问题属于每个预设问题标签的概率分布,从而,可以先对上述概率分布进行归一化处理,进而,选取概率最大的预设问题标签为上述样本问题所关联的第四问题标签。

[0133] 在步骤S1103中,根据第三问题标签和第四问题标签之间的一致性比对结果,对初始训练样本进行数据清洗,得到清洗后训练样本。

[0134] 本步骤中,可以将上述第三问题标签和上述第四问题标签进行一致性比对,若上述第三问题标签与上述第四问题标签一致,则可以将上述样本问题保留在上述初始训练样本中。若第三问题标签与第四问题标签不一致,则可以将上述样本问题从上述初始训练样本中剔除,以对上述初始训练样本进行数据清洗,得到清洗后训练样本。

[0135] 在得到清洗后训练样本之后,接着参考图2,在步骤S204中,将清洗后训练样本作为训练样本,并根据训练样本对初始领域检测模型进行迭代训练,得到更新后的领域检测模型。

[0136] 本步骤中,可以将上述清洗后训练样本作为训练样本,对上述初始领域检测模型进行迭代训练,得到更新后的领域检测模型,以提升模型质量。

[0137] 可选的,在得到更新后的领域检测模型之后,还可以多次重复上述步骤S203-步骤S204,以对上述更新后的领域检测模型进行多次更新,可以根据实际情况自行设定,本公开对此不作特殊限定。

[0138] 通过对初始训练样本进行多次数据清洗,以及,对初始领域检测模型进行多次迭

代更新,能够使得训练样本的数据质量不断提升,从而提升模型的应答质量。

[0139] 在得到更新后的领域检测模型之后,可以接着参考图1,在步骤S110中,采集语料数据,并从语料数据中提取备选问题。

[0140] 本步骤中,语料数据可以是线上收集到的日志数据,即用户在使用智能应答系统的过程中所输入的问题数据,这些问题数据可以是没有问题标签的数据。通过采集语料数据,使得本公开可以利用新的线上数据对训练样本进行扩充,相比于现有技术中仅基于已有数据进行扩充的方案而言,本公开能够在训练样本中增添新的数据,从而,提升模型的泛化能力。

[0141] 在采集到语料数据之后,可以进入步骤S120中,通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签。

[0142] 本步骤中,可以参考图12,图12示出本公开实施例中如何通过更新后的领域检测模型基于第一识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第一问题标签的流程示意图,包含步骤S1201-步骤S1204:

[0143] 在步骤S1201中,通过更新后的领域检测模型提取备选问题对应的隐特征向量。

[0144] 本步骤中,上述更新后的领域检测模型也可以包含n个网络层,从而,参考上述步骤S601-步骤S603的相关解释可知,可以对n个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量,进而,计算第n个网络层对应的输出向量与每个网络层对应的输出向量之间的余弦相似度,得到n个余弦相似度,之后,可以将n个相似度作为n个网络层对应的输出向量的加权系数,对n个网络层对应的n个输出向量进行加权处理,得到备选问题对应的隐特征向量。

[0145] 在步骤S1202中,获取每个预设问题标签对应的隐特征向量。

[0146] 本步骤中,鉴于上述步骤S502中已经计算并存储了每个预设问题标签对应的隐特征向量,从而,该步骤可以直接调用预先存储的数据,获得每个预设问题标签对应的隐特征向量。

[0147] 在步骤S1203中,计算备选问题对应的隐特征向量与每个预设问题标签对应的隐特征向量之间的第一向量距离。

[0148] 本步骤中,可以计算备选问题对应的隐特征向量与每个预设问题标签对应的隐特征向量之间的第一向量距离,示例性的,第一向量距离可以是马氏距离。

[0149] 在步骤S1204中,根据多个第一向量距离,确定备选问题所关联的第一问题标签。

[0150] 本步骤中,参考图13,图13示出本公开实施例中如何根据多个第一向量距离,确定备选问题所关联的第一问题标签的流程示意图,包含步骤S1301-步骤S1304:

[0151] 在步骤S1301中,从多个第一向量距离中筛选出数值最小的第一向量距离,并将数值最小的第一向量距离对应的预设问题标签确定为备选问题所关联的候选第一问题标签。

[0152] 本步骤中,可以从多个上述第一向量距离中筛选出数值最小的第一向量距离,并将数值最小的第一向量距离对应的预设问题标签确定为备选问题所关联的候选第一问题标签。

[0153] 在步骤S1302中,查找与候选第一问题标签对应的指定距离阈值。

[0154] 本步骤中,鉴于上述步骤S802中已经计算并存储了每个预设问题标签对应的指定

距离阈值,从而,本步骤中,可以直接查找与上述候选第一问题标签对应的指定距离阈值。

[0155] 在步骤S1303中,若数值最小的第一向量距离大于指定距离阈值,确定备选问题与候选第一问题标签之间不具备关联关系。

[0156] 本步骤中,若上述数值最小的第一向量距离大于指定距离阈值,确定备选问题与候选第一问题标签之间不具备关联关系,从而,可以直接舍弃该备选问题,无需利用该备选问题扩充训练样本。

[0157] 在步骤S1304中,若数值最小的第一向量距离不大于指定距离阈值,将候选第一问题标签确定为备选问题所关联的第一问题标签。

[0158] 本步骤中,若上述数值最小的第一向量距离不大于上述指定距离阈值,则可以将上述第一问题标签确定为备选问题所关联的第一问题标签。

[0159] 接着参考图1,在步骤S130中,通过更新后的领域检测模型基于第二识别逻辑识别备选问题与多个预设问题标签之间的关联关系,获得备选问题所关联的第二问题标签。

[0160] 本步骤中,鉴于在步骤S1201中已经提取出了每个备选问题对应的隐特征向量,而隐特征向量中包含了CLS向量,从而,更新后的领域检测模型可以从上述备选问题对应的隐特征向量中抽取CLS向量(CLS向量用于表征备选问题与每个预设问题标签的相关性),根据CLS向量,从多个预设问题标签中确定备选问题所关联的第二问题标签。

[0161] 在步骤S140中,若第一问题标签与第二问题标签一致,利用备选问题扩充更新后的领域检测模型对应的训练样本。

[0162] 本步骤中,如果上述步骤S1304中确定出来的第一问题标签与上述步骤S130中确定出来的第二问题标签一致,则可以利用备选问题扩充更新后的领域检测模型对应的训练样本(示例性的,可以将上述备选问题归类至上述第一问题标签下),否则,舍弃上述备选问题,跳转至下一个备选问题,并判断是否利用下一个备选问题扩充更新后的领域检测模型对应的训练样本。

[0163] 参考图14,图14示出本公开实施例中训练样本扩充方法的整体示意图,如图14所示:

[0164] 首先,可以获取初始训练样本;

[0165] 其次,利用初始训练样本训练得到初始领域检测模型,利用初始领域检测模型清洗上述初始训练样本,并利用清洗后的初始训练样本对初始领域检测模型进行迭代训练,得到更新后的领域检测模型;

[0166] 再次,利用更新后的领域检测模型对线上日志数据进行筛选;

[0167] 最后,利用更新后的领域检测模型输出用于扩充训练样本的增强数据。

[0168] 整体而言,一方面,本公开通过训练初始领域检测模型,再通过初始领域检测模型对初始训练样本进行清洗,利用清洗后的初始训练样本对初始领域检测模型进行迭代训练,得到更新后的领域检测模型,能够在多次迭代之后获得检测效果较好的领域检测模型,在此基础上,能够过滤掉初始训练样本中的错误以及混淆数据,提升初始训练样本的数据质量,从而,保证智能问答系统的问答效果;

[0169] 另一方面,本公开通过利用更新后的领域检测模型基于两种识别逻辑对线上日志数据中包含的备选问题与上述训练样本中包含的预设问题标签之间的关联关系进行识别,以将备选问题划分至对应的知识点或提问意图,获得备选问题所关联的第一问题标签和第

二问题标签,通过第一问题标签和第二问题标签之间的对比验证,能够在准确理清备选问题与多个预设问题标签之间的关联关系之后,再基于备选问题扩充训练样本,从而,解决了相关技术中在样本扩充时人工划分关联关系所导致的出现混淆或错误数据的问题,提升扩充之后的训练样本的数据质量,从而,进一步提升智能问答系统的问答效果。

[0170] 本公开还提供了一种训练样本扩充装置,图15示出本公开示例性实施例中训练样本扩充装置的结构示意图;如图15所示,训练样本扩充装置1500可以包括数据采集模块1510、第一识别模块1520、第二识别模块1530和样本扩充模块1540。其中:

[0171] 数据采集模块1510,用于采集语料数据,并从所述语料数据中提取备选问题;

[0172] 第一识别模块1520,用于通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签;

[0173] 第二识别模块1530,用于通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签;

[0174] 样本扩充模块1540,用于若所述第一问题标签与所述第二问题标签一致,利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

[0175] 在本公开的示例性实施例中,所述预设问题标签包括所述知识点标签或提问意图标签。

[0176] 在本公开的示例性实施例中,第一识别模块1520通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签,包括:通过更新后的领域检测模型提取所述备选问题对应的隐特征向量;获取每个预设问题标签对应的隐特征向量;计算所述备选问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第一向量距离;根据多个所述第一向量距离,确定所述备选问题所关联的第一问题标签。

[0177] 在本公开的示例性实施例中,所述更新后的领域检测模型包括 n 个网络层, n 为大于1的整数;第一识别模块1520通过更新后的领域检测模型提取所述备选问题对应的隐特征向量,包括:对所述 n 个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;计算第 n 个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到 n 个相似度;将所述 n 个相似度作为所述 n 个网络层对应的输出向量的加权系数,对 n 个所述网络层对应的 n 个输出向量进行加权处理,得到所述备选问题对应的隐特征向量。

[0178] 在本公开的示例性实施例中,第一识别模块1520根据多个所述第一向量距离,确定所述备选问题所关联的第一问题标签,包括:从多个所述第一向量距离中筛选出数值最小的第一向量距离,并将所述数值最小的第一向量距离对应的预设问题标签确定为所述备选问题所关联的候选第一问题标签;查找与所述候选第一问题标签对应的指定距离阈值;若所述数值最小的第一向量距离大于所述指定距离阈值,确定所述备选问题与所述候选第一问题标签之间不具备关联关系;若所述数值最小的第一向量距离不大于所述指定距离阈值,将所述候选第一问题标签确定为所述备选问题所关联的第一问题标签。

[0179] 在本公开的示例性实施例中,在确定所述样本问题与所述候选第一问题标签之间

不具备关联关系之后,第一识别模块1520被配置为:

[0180] 舍弃所述备选问题。

[0181] 在本公开的示例性实施例中,第二识别模块1530通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签,包括:从所述备选问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述备选问题与每个所述预设问题标签的相关性;根据所述CLS向量,从所述多个预设问题标签中确定所述备选问题所关联的第二问题标签。

[0182] 在本公开的示例性实施例中,样本扩充模块1540,被配置为:

[0183] 获取初始训练样本,所述初始训练样本包括多个预设问题标签,每个所述预设问题标签下关联有多个样本问题;根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型;利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本;将所述清洗后训练样本作为所述训练样本,并根据所述训练样本对所述初始领域检测模型进行迭代训练,得到所述更新后的领域检测模型。

[0184] 在本公开的示例性实施例中,样本扩充模块1540根据所述初始训练样本对待训练的领域检测模型进行迭代训练,得到初始领域检测模型,包括:将所述初始训练样本输入至所述待训练的领域检测模型中,得到所述待训练的领域检测模型输出的每个所述样本问题所关联的预测问题标签;根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值;根据所述损失值,对所述待训练的领域检测模型进行迭代训练,得到所述初始领域检测模型。

[0185] 在本公开的示例性实施例中,样本扩充模块1540根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,以及,所述多个样本问题中任意两个样本问题之间的差异程度,确定所述待训练的领域检测模型的损失值,包括:根据所述预测问题标签与所述样本问题所关联的预设问题标签之间的差异程度,计算第一类损失值;以及,根据所述多个样本问题中任意两个样本问题之间的差异程度,计算第二类损失值;根据所述第一类损失值和所述第二类损失值,确定所述待训练的领域检测模型的损失值。

[0186] 在本公开的示例性实施例中,样本扩充模块1540利用所述初始领域检测模型对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量;根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量;根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签;根据所述第三问题标签与所述目标样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

[0187] 在本公开的示例性实施例中,所述初始领域检测模型包括n个网络层,n为大于1的整数;样本扩充模块1540利用所述初始领域检测模型获取所述每个样本问题对应的隐特征向量,包括:对所述n个网络层中每个网络层的输出特征进行池化处理,得到每个网络层对应的输出向量;计算第n个网络层对应的输出向量与所述每个网络层对应的输出向量之间的相似度,得到n个相似度;将所述n个相似度作为所述n个网络层对应的输出向量的加权系

数,对n个所述网络层对应的n个输出向量进行加权处理,得到所述每个样本问题对应的隐特征向量。

[0188] 在本公开的示例性实施例中,样本扩充模块1540根据所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量,确定所述每个预设问题标签对应的隐特征向量,包括:对所述每个预设问题标签下所关联的目标样本问题对应的隐特征向量进行求均值处理,得到所述每个预设问题标签对应的隐特征向量。

[0189] 在本公开的示例性实施例中,样本扩充模块1540根据所述样本问题对应的隐特征向量与所述每个预设问题标签对应的隐特征向量之间的第二向量距离,确定所述每个样本问题所关联的第三问题标签,包括:从多个所述第二向量距离中筛选出数值最小的第二向量距离,并将所述数值最小的第二向量距离对应的预设问题标签确定为所述样本问题所关联的候选第三问题标签;计算所述候选第三问题标签对应的指定距离阈值,存储所述候选第三问题标签与所述指定距离阈值之间的对应关系;若所述数值最小的第二向量距离大于所述指定距离阈值,确定所述样本问题与所述候选第三问题标签之间不具备关联关系;若所述数值最小的第二向量距离不大于所述指定距离阈值,将所述候选第三问题标签确定为所述样本问题所关联的第三问题标签。

[0190] 在本公开的示例性实施例中,样本扩充模块1540计算所述候选第三问题标签对应的指定距离阈值,包括:计算所述候选第三问题标签对应的隐特征向量与其他所述预设问题标签对应的隐特征向量之间的多个第三向量距离;以及,计算所述候选第三问题标签所关联的多个样本问题对应的隐特征向量与所述候选第三问题标签对应的隐特征向量之间的多个第四向量距离;根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值。

[0191] 在本公开的示例性实施例中,样本扩充模块1540根据所述多个第三向量距离和所述多个第四向量距离,确定所述候选第三问题标签对应的指定距离阈值,包括:从所述多个第三向量距离中筛选数值最小的第三向量距离,从所述多个第四向量距离中筛选出数值最大的第四向量距离;获取所述数值最小的第三向量距离的目标关联值;将所述目标关联值与所述数值最大的第四向量距离之间的较大值确定为所述候选第三问题标签对应的指定距离阈值。

[0192] 在本公开的示例性实施例中,样本扩充模块1540根据所述第三问题标签与所述样本问题所关联的预设问题标签之间的一致性对比结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本,包括:若所述第三问题标签与所述样本问题所关联的预设问题标签不一致,则将所述样本问题从所述初始训练样本中剔除,以得到所述清洗后训练样本;若所述第三问题标签与所述样本问题所关联的预设问题标签一致,则根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签;根据所述第三问题标签和所述第四问题标签之间的一致性比对结果,对所述初始训练样本进行数据清洗,得到清洗后训练样本。

[0193] 在本公开的示例性实施例中,样本扩充模块1540根据所述每个样本问题对应的隐特征向量,识别所述样本问题所关联的第四问题标签,包括:从所述样本问题对应的隐特征向量中抽取CLS向量,所述CLS向量用于表征所述样本问题与每个所述预设问题标签的相关性;根据所述CLS向量,从所述多个预设问题标签中确定所述样本问题所关联的第四问题标

签。

[0194] 上述训练样本扩充装置中各模块的具体细节已经在对应的训练样本扩充方法中进行了详细的描述,因此此处不再赘述。

[0195] 应当注意,尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元,但是这种划分并非强制性的。实际上,根据本公开的实施方式,上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之,上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0196] 此外,尽管在附图中以特定顺序描述了本公开中方法的各个步骤,但是,这并非要求或者暗示必须按照该特定顺序来执行这些步骤,或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的,可以省略某些步骤,将多个步骤合并为一个步骤执行,以及/或者将一个步骤分解为多个步骤执行等。

[0197] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者网络设备等)执行根据本公开实施方式的方法。

[0198] 本申请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可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。

[0199] 计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0200] 计算机可读存储介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读存储介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0201] 计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被一个该电子设备执行时,使得该电子设备实现如上述实施例中所述的方法。

[0202] 此外,在本公开实施例中还提供了一种能够实现上述方法的电子设备。

[0203] 所属技术领域的技术人员能够理解,本公开的各个方面可以实现为系统、方法或程序产品。因此,本公开的各个方面可以具体实现为以下形式,即:完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等),或硬件和软件方面结合的实施方式,这里可以统称为“电路”、“模块”或“系统”。

[0204] 下面参照图16来描述根据本公开的这种实施方式的电子设备1600。图16显示的电子设备1600仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0205] 如图16所示,电子设备1600以通用计算设备的形式表现。电子设备1600的组件可

以包括但不限于：上述至少一个处理单元1610、上述至少一个存储单元1620、连接不同系统组件(包括存储单元1620和处理单元1610)的总线1630以及显示单元1640。

[0206] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元1610执行,使得所述处理单元1610执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。例如,所述处理单元1610可以执行如图1中所示的:步骤S110,采集语料数据,并从所述语料数据中提取备选问题;步骤S120,通过更新后的领域检测模型基于第一识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第一问题标签;步骤S130,通过所述更新后的领域检测模型基于第二识别逻辑识别所述备选问题与多个预设问题标签之间的关联关系,获得所述备选问题所关联的第二问题标签;步骤S140,若所述第一问题标签与所述第二问题标签一致,利用所述备选问题扩充所述更新后的领域检测模型对应的训练样本。

[0207] 存储单元1620可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM) 16201和/或高速缓存存储单元16202,还可以进一步包括只读存储单元(ROM) 16203。

[0208] 存储单元1620还可以包括具有一组(至少一个)程序模块16205的程序/实用工具16204,这样的程序模块16205包括但不限于:操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0209] 总线1630可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0210] 电子设备1600也可以与一个或多个外部设备1700(例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与该电子设备1600交互的设备通信,和/或与使得该电子设备1600能与一个或多个其它计算设备进行通信的任何设备(例如路由器、调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口1650进行。并且,电子设备1600还可以通过网络适配器1660与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器1660通过总线1630与电子设备1600的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备1600使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0211] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其他实施例。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由权利要求指出。

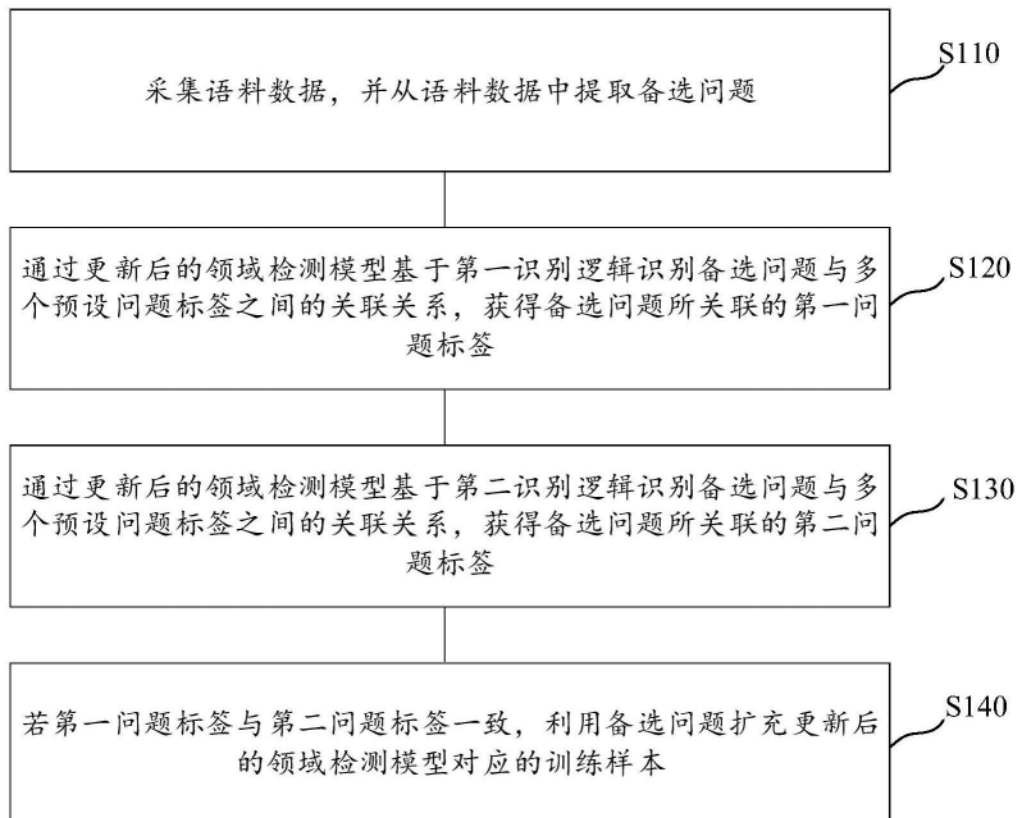


图1

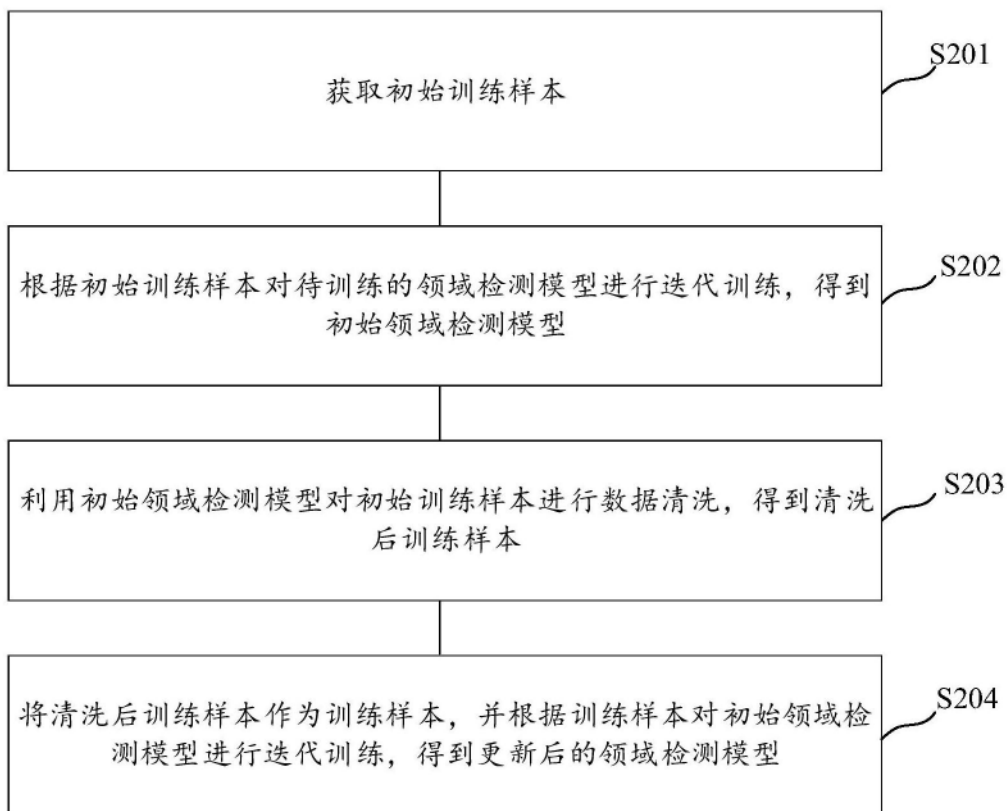


图2

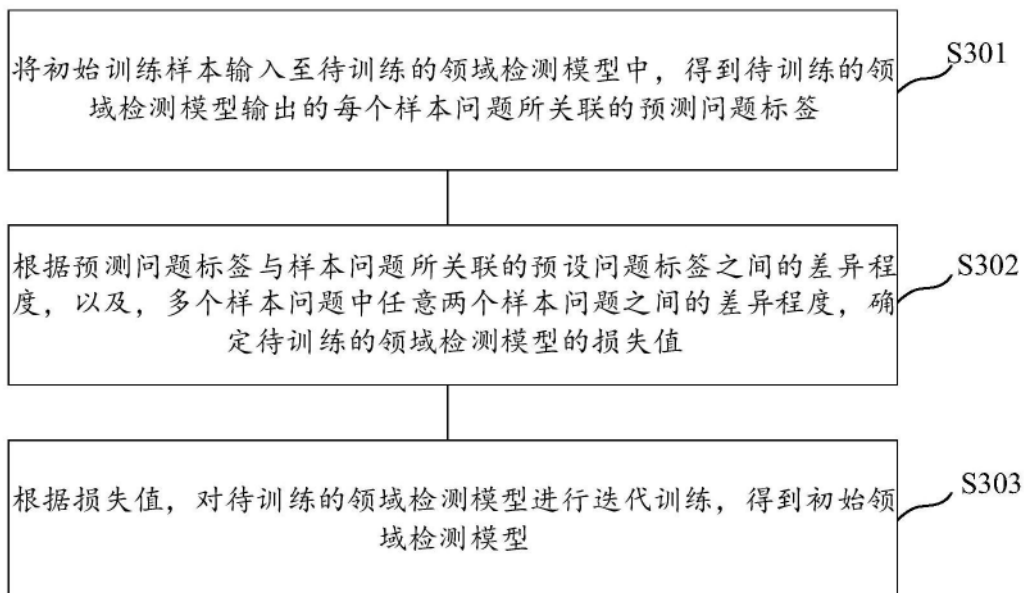


图3

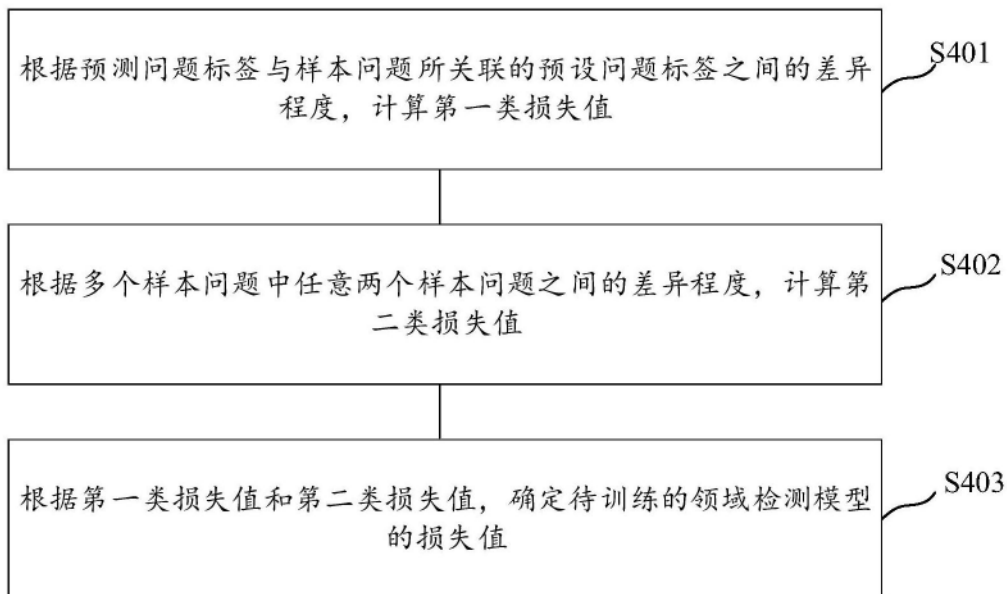


图4

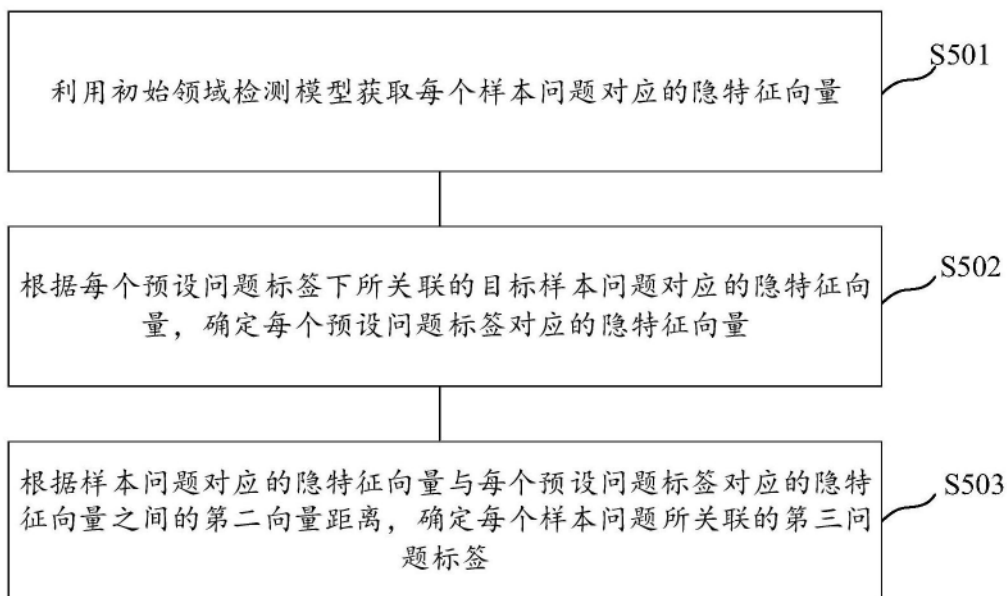


图5

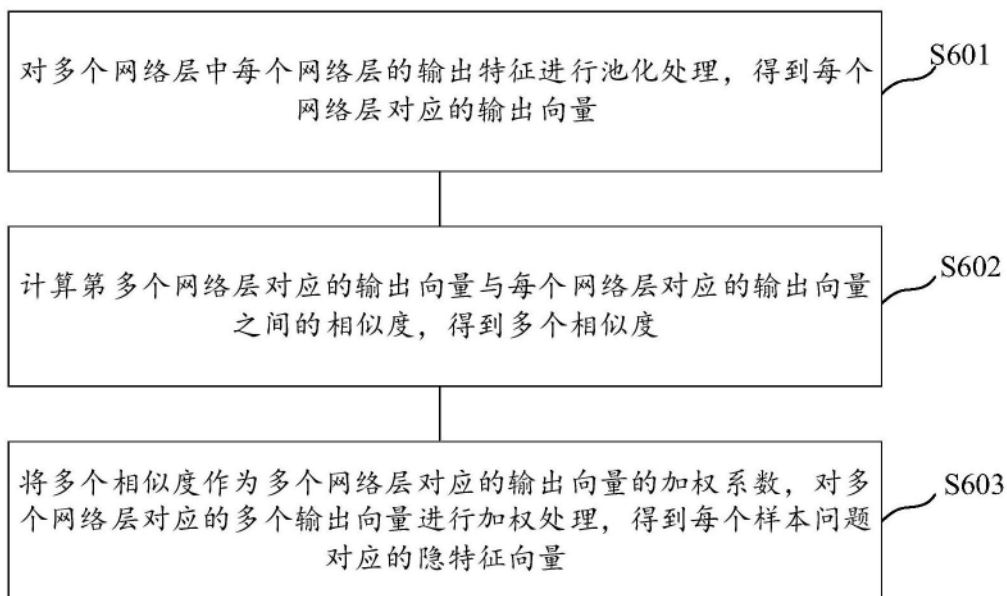


图6

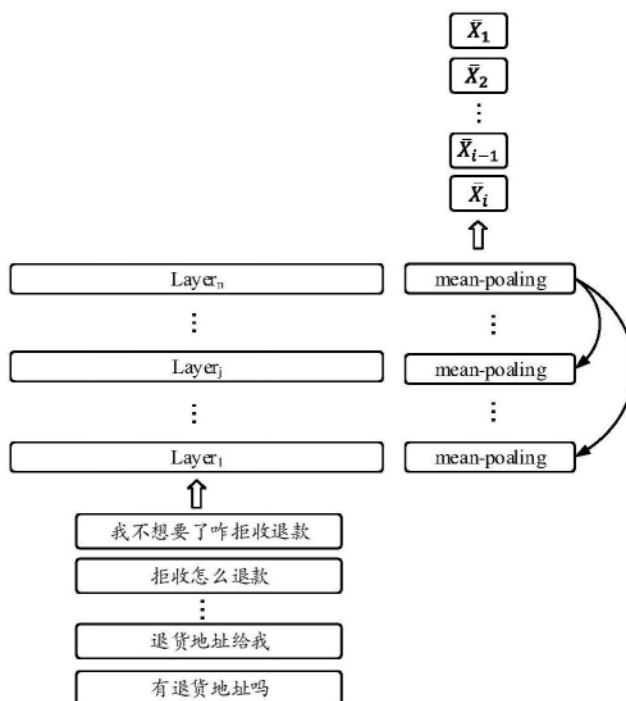


图7

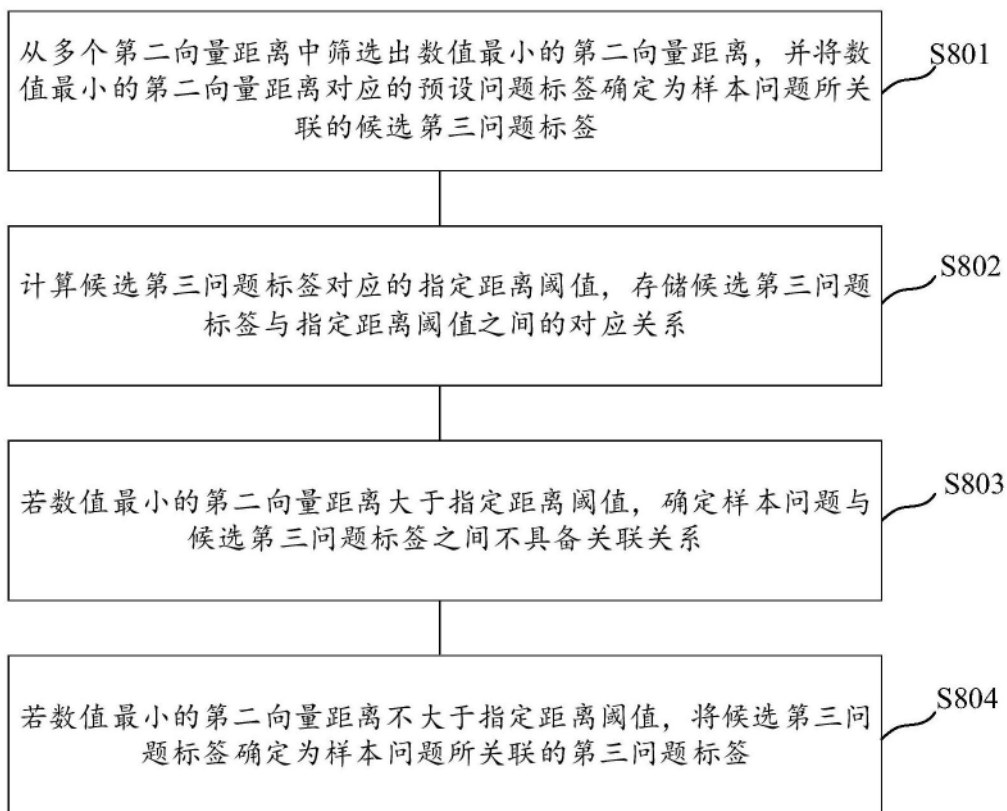


图8

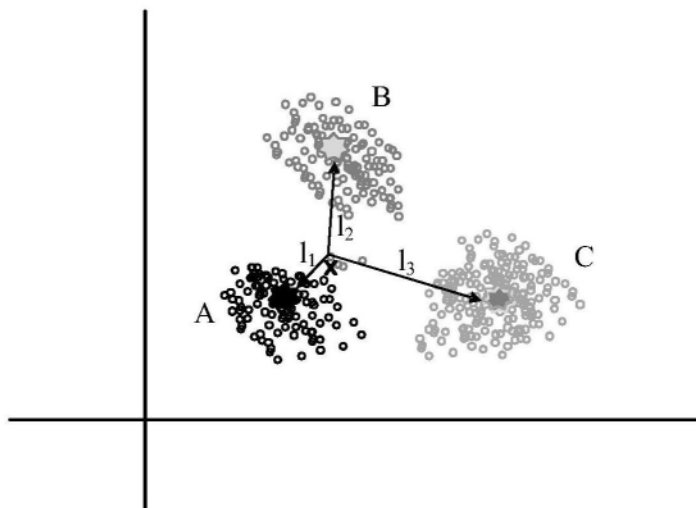


图9

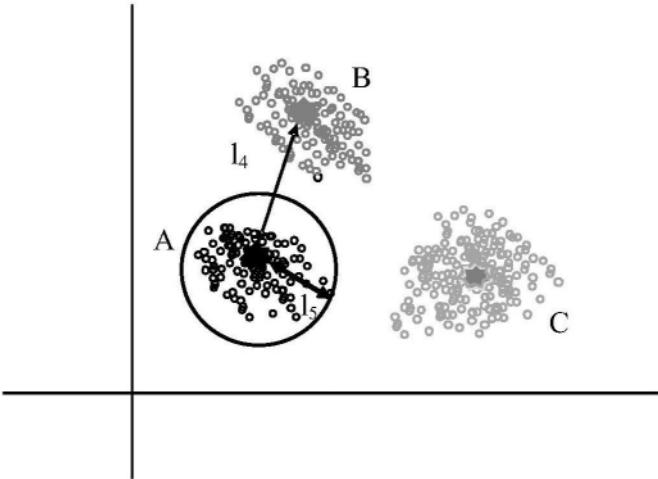


图10

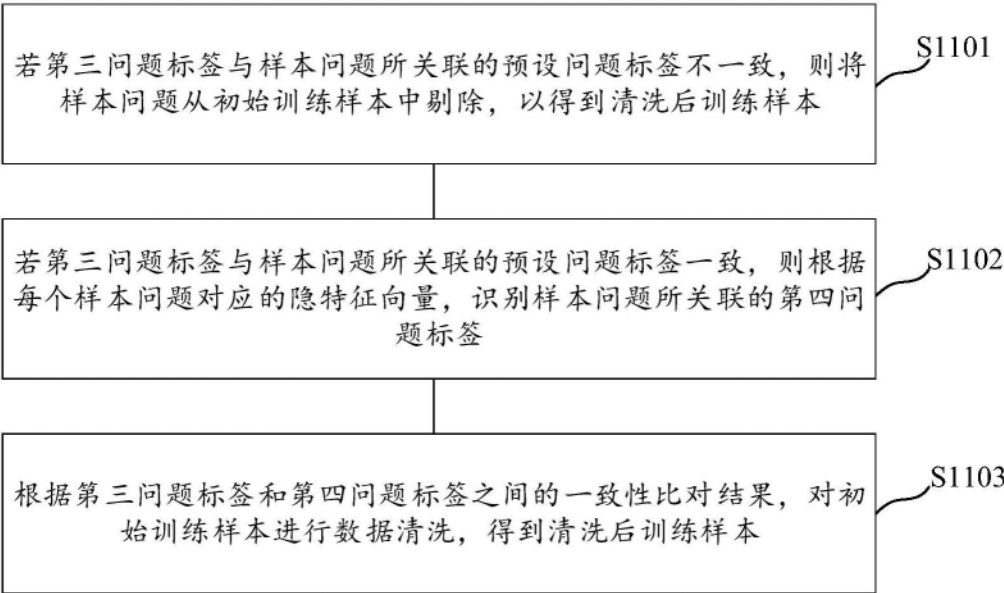


图11



图12

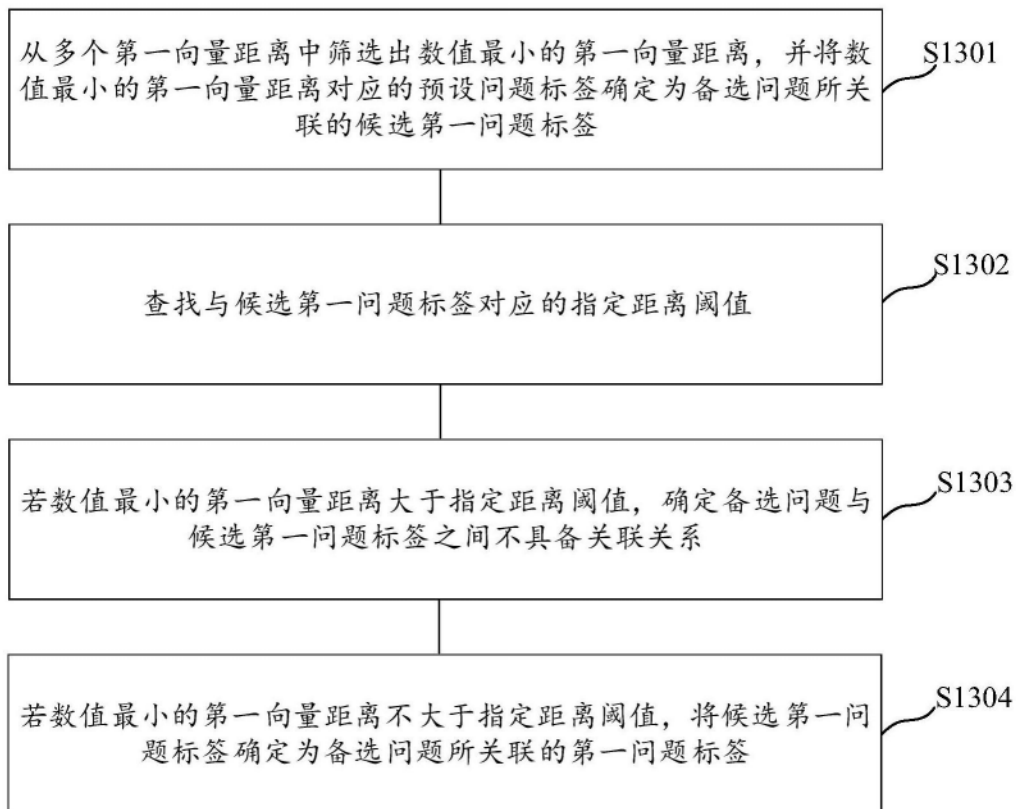


图13

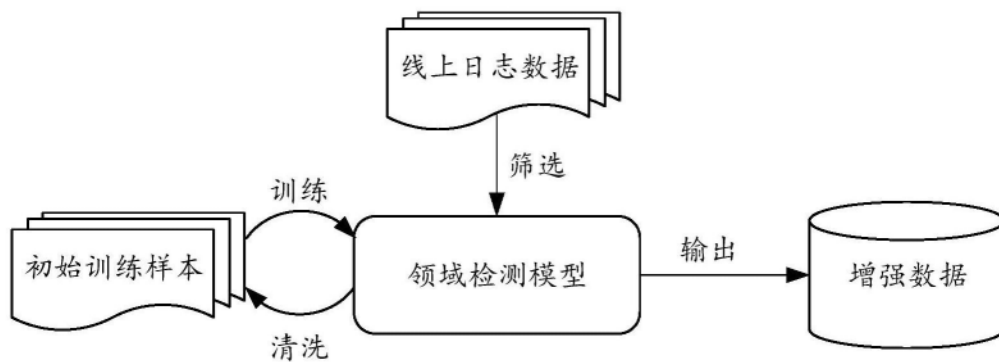


图14

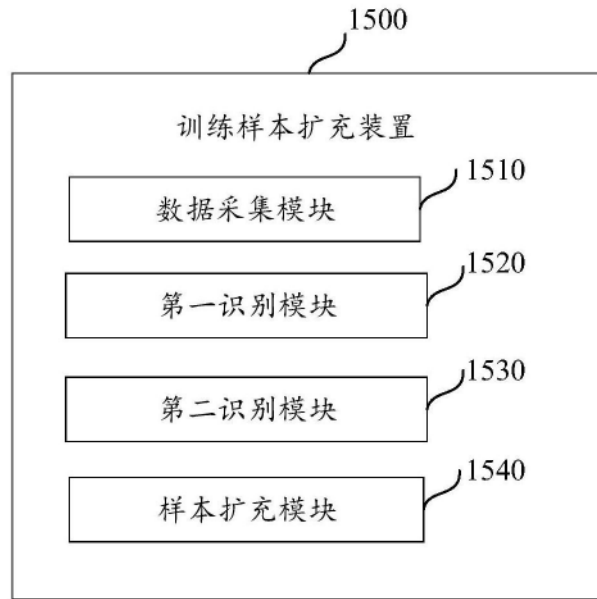


图15

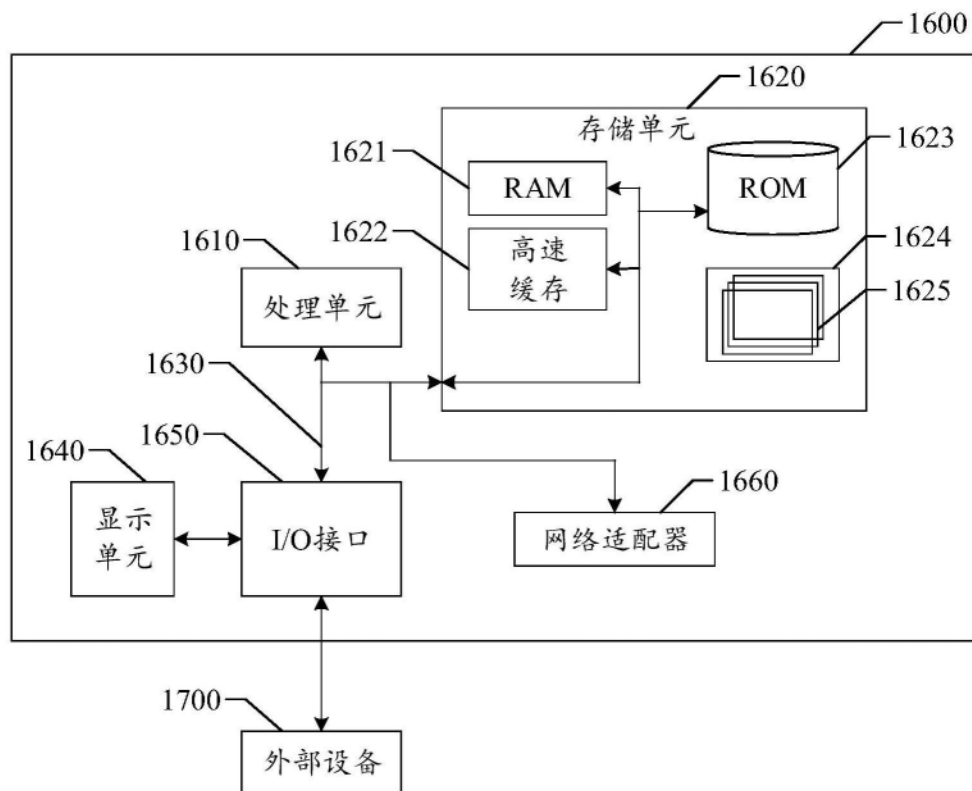


图16