

(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
2020 年 7 月 16 日 (16.07.2020)

(10) 国际公布号
WO 2020/143301 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01) **G06F 16/35** (2019.01)

(21) 国际申请号: PCT/CN2019/117090

(22) 国际申请日: 2019 年 11 月 11 日 (11.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910022265.4 2019 年 1 月 10 日 (10.01.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 许开河(XU, Kaihe); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 杜尧鑫(DU, Yaoxin); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 王少军(WANG, Shaojun); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 北京汇思诚业知识产权代理有限公司 (UNI-INTEL PATENT AND TRADEMARK LAW FIRM); 中国北京市朝阳区建国门外大街永安东里甲 3 号通用国际中心 A 座 3 层, Beijing 100022 (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,

(54) **Title:** TRAINING SAMPLE VALIDITY DETECTION METHOD, COMPUTER DEVICE, AND COMPUTER NON-VOLATILE STORAGE MEDIUM

(54) 发明名称: 一种训练样本有效性检测方法、计算机设备及计算机非易失性存储介质

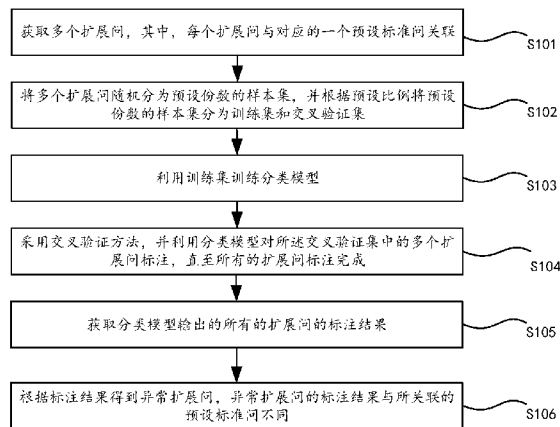


图 1

S101 Acquire multiple extended questions, wherein each extended question is associated with a corresponding preset standard question

S102 Randomly divide the multiple expansion questions into preset copies of sample sets and divide the preset copies of sample sets into a training set and a cross-validation set according to a preset ratio

S103 Train a classification model by using the training set

S104 Use the classification model to label the multiple extended questions in the cross-validation set by adopting a cross-validation method until all the extended questions are labeled

S105 Acquire the labeling results of all the extended questions output by the classification model

S106 Obtain abnormal extended questions according to the labeling results, the labeling results of the abnormal extended questions being different from the associated preset standard questions

(57) **Abstract:** A training sample validity detection method, a computer device, and a computer non-volatile storage medium, relating to the technical field of artificial intelligence. The method comprises: acquiring multiple extended questions, wherein each extended question is associated with a corresponding preset standard question (S101); randomly dividing the multiple expansion questions into preset copies of sample sets and dividing the preset copies of sample sets into a training set and a cross-validation set according to a preset ratio (S102); training a classification model by using the training set (S103); using the classification model to label the multiple

BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))。

extended questions in the cross-validation set by adopting a cross-validation method until all the extended questions are labeled (S104); acquiring the labeling results of all the extended questions output by the classification model (S105); and obtaining abnormal extended questions according to the labeling results, the labeling results of the abnormal extended questions being different from the associated preset standard questions (S106). The present application can solve the problem of low efficiency of training sample validity detection in the prior art.

(57) 摘要: 一种训练样本有效性检测方法、计算机设备及计算机非易失性存储介质, 涉及人工智能技术领域, 方法包括: 获取多个扩展问, 其中, 每个扩展问与对应的一个预设标准问关联(S101); 将多个扩展问随机分为预设份数的样本集, 并根据预设比例将预设份数的样本集分为训练集和交叉验证集(S102); 利用训练集训练分类模型(S103); 采用交叉验证方法, 并利用分类模型对交叉验证集中的多个扩展问标注, 直至所有的扩展问标注完成(S104); 获取分类模型输出的所有的扩展问的标注结果(S105); 根据标注结果得到异常扩展问, 异常扩展问的标注结果与所关联的预设标准问不同(S106)。能够解决现有技术中训练样本有效性检测效率低的问题。

一种训练样本有效性检测方法、计算机设备及计算机非易失性存储介质

本申请要求于 2019 年 01 月 10 日提交中国专利局、申请号为 201910022265.4、申请名称为“一种训练样本有效性检测方法及装置”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】

本申请涉及人工智能技术领域，尤其涉及一种训练样本有效性检测方法、计算机设备及计算机非易失性存储介质。

【背景技术】

目前，客服机器人发展趋于成熟，根据所需要实现的业务场景，需要大量的训练语料对基于机器学习智能应答模型进行训练，然而，现实中训练语料往往质量参差不齐，不好的语料会导致应答模型的回答效果产生负面影响，因此如何提高训练样本的有效性检测效率成为目前亟待解决的问题。

【申请内容】

有鉴于此，本申请实施例提供了一种训练样本有效性检测方法、计算机设备及计算机非易失性存储介质，用以解决现有技术中训练样本的有效性检测效率低的问题。

为了实现上述目的，根据本申请的一个方面，提供了一种训练样本有效性检测方法，所述方法包括：

步骤S101，获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

步骤S102，将多个所述扩展问随机分为预设份数的样本集，并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集；

步骤S103, 利用所述训练集训练分类模型;

步骤S104, 采用交叉验证方法, 并利用所述分类模型对所述交叉验证集中的多个扩展问标注, 直至所有的所述扩展问标注完成;

步骤S105, 获取所述分类模型输出的所有的所述扩展问的标注结果;

步骤S106, 根据所述标注结果得到异常扩展问, 所述异常扩展问的标注结果与所关联的预设标准问不同。

为了实现上述目的, 根据本申请的一个方面, 提供了一种训练样本有效性检测装置, 所述装置包括:

第一获取单元, 用于获取多个扩展问, 其中, 每个所述扩展问与对应的一个预设标准问关联; 制作单元, 用于将多个所述扩展问随机分为预设份数的样本集, 并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集; 训练单元, 用于利用所述训练集训练分类模型; 标注单元, 用于采用交叉验证方法, 并利用所述分类模型对所述交叉验证集中的多个扩展问标注, 直至所有的所述扩展问标注完成; 第二获取单元, 用于获取所述分类模型输出的所有的所述扩展问的标注结果; 第一处理单元, 用于根据所述标注结果得到异常扩展问, 所述异常扩展问的标注结果与所关联的预设标准问不同。

为了实现上述目的, 根据本申请的一个方面, 提供了一种计算机非易失性存储介质, 所述存储介质包括存储的程序, 在所述程序运行时控制所述存储介质所在设备执行上述的训练样本有效性检测方法的步骤。

为了实现上述目的, 根据本申请的一个方面, 提供了一种计算机设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序, 所述处理器执行所述计算机程序时实现上述的训练样本有效性检测方法的步骤。

在本方案中, 利用交叉验证方法和分类模型来对扩展问进行标注, 从而根据标注结果与扩展问关联的预设标准问进行匹配, 得到异常扩展问, 快速筛选出异常扩展问, 提高了训练样本的有效性检测效率, 从而避免一些质量低的扩展问来训练智能应答模型。

【附图说明】

为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其它的附图。

图 1 是根据本申请实施例的一种训练样本有效性检测方法的流程图;

图 2 是根据本申请实施例的一种训练样本有效性检测装置的示意图;

图 3 是根据本申请实施例的一种计算机设备的示意图。

【具体实施方式】

为了更好的理解本申请的技术方案,下面结合附图对本申请实施例进行详细描述。

应当明确,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本申请保护的范围。

在本申请实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本申请。在本申请实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

应当理解,本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A 和/或 B,可以表示:单独存在 A,同时存在 A 和 B,单独存在 B 这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。

应当理解,尽管在本申请实施例中可能采用术语第一、第二、第三等来描述终端,但这些终端不应限于这些术语。这些术语仅用来将终端

彼此区分开。例如，在不脱离本申请实施例范围的情况下，第一获取单元也可以被称为第二获取单元，类似地，第二获取单元也可以被称为第一获取单元。

取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测（陈述的条件或事件）”可以被解释成为“当确定时”或“响应于确定”或“当检测（陈述的条件或事件）时”或“响应于检测（陈述的条件或事件）”。

图 1 是根据本申请实施例的一种训练样本有效性检测方法的流程图，如图 1 所示，该方法包括：

步骤S101，获取多个扩展问，其中，每个扩展问与对应的一个预设标准问关联。

步骤S102，将多个扩展问随机分为预设份数的样本集，并根据预设比例将预设份数的样本集分为训练集和交叉验证集。

步骤S103，利用训练集训练分类模型。

步骤S104，采用交叉验证方法，并利用分类模型对交叉验证集中的多个扩展问标注，直至所有的扩展问标注完成。

步骤S105，获取分类模型输出的所有的扩展问的标注结果。

步骤S106，根据标注结果得到异常扩展问，异常扩展问的标注结果与所关联的预设标准问不同。

在本方案中，利用交叉验证方法和分类模型来对扩展问进行标注，从而根据标注结果与扩展问关联的预设标准问进行匹配，得到异常扩展问，快速筛选出异常扩展问，提高了训练样本的有效性检测效率，从而避免一些质量低的扩展问来训练智能应答模型。

例如，预设标准问例如为“信用卡逾期怎么办？”，它的扩展问 1 为“我的信用卡逾期了，需要怎么处理”，扩展问 2 为“信用卡逾期可以不还吗”，扩展问 3 为“过了信用卡还款日，怎么办”。可以理解地，扩展问 1、扩展问 2、扩展问 3 所关联的预设标准问皆为“信用卡逾期怎么办？”

在本实施例中，训练集用于对分类模型进行训练样本集，为了使训练后的分类模型能够基于用户输入的问题识别出用户可能期望咨询的标准问。训练集中的扩展问为待分类的数据，其类别即为其关联的预设标准问。具体地，分类模型包括输入层、输出层及融合层。其中，输入层包括词语特征输入层和问题特征输入层。每个扩展问所关联的预设标准问作为问题特征，问题特征输入层用于获取输入的扩展问关联的预设标准问的特征，词语特征输入层用于获取扩展问分词处理后得到的词汇的特征。分类模型在经过训练后，能够根据学习到的词汇特性及问题特征来对一个新的扩展问进行分类，使得新的扩展问尽可能地分类至一个准确的标准问下。

可选地，交叉验证方法为十折交叉验证方法或五折交叉验证方法中的任意一种。在一种实施方式中，采用五折交叉验证方法，具体地，将多个扩展问随机分成10份，每次取其中2份作为交叉验证集，其余8份作为训练集。训练时，先用其中的8份对初始化后的分类模型进行训练，然后对2份交叉验证集进行分类标注，以此重复训练及验证过程5次，每次选取的交叉验证集不同，直至所有的扩展问都被标注一遍。可以理解地，根据分类模型的分类结果对交叉验证集中的扩展问进行标注，例如将扩展问4“我的信用卡要逾期了，我该怎么处理”分类至预设标准问“信用卡逾期怎么办”下，那么扩展问4的关联标准问就是分类模型的分类结果。

在另一种实施方式中，将每个预设标准问所关联的多个扩展问随机分成10份，每次从每个预设标准问所关联的10份扩展问中抽取1份作为交叉验证集，其余9份作为训练集。例如：预设标准问有10个，分别为A1，A2...A10，每个预设标准问关联多个扩展问，因此将每个预设标准问所关联的多个扩展问随机分成10份，可以保证交叉验证时每个预设标准问都能得到有效的训练。

可选地，在步骤S105之后，方法还包括：

重复操作M次的步骤S102至步骤S104，其中，M为大于1的自然数，每次重复操作前对分类模型进行初始化设置；获取分类模型的输出的所有的扩展问的M次标注结果；当标注结果与预设标准问相同时，扩展问的当次

计分为第一预设分值；或，当标注结果与预设标准问不同时，扩展问的当次计分为第二预设分值；计算各个扩展问的第一累计得分；筛选出第一累计得分小于第一预设阈值的扩展问；将筛选出的扩展问作为异常扩展问，并输出。

具体地，分类模型标注M次后，对M次的所有扩展问的标注结果进行统计，扩展问的得分越低，其关联的预设标准问是错误的可能性越高。例如：当标注结果与预设标准问相同时，扩展问的得分为1；当标注结果与预设标准问不同时，扩展问的得分为-1，重复标注10次，其中6次得分为1，4次得分为-1，则扩展问的得分为10次标注结果的加和平均值，为0.2分。当第一预设阈值为0.1分时，则表示多个扩展问的分值小于0.1分的皆为异常扩展问。

可选地，所述分类模型可以采用多种机器学习算法，可以是朴素贝叶斯算法、最近邻法则、线性判别分析、支持向量机或决策树算法。在重复操作M次的步骤S102至步骤S104中，所述分类模型所采用的机器学习算法可以相同，也可以不同。采用不同的机器学习算法，能够减小标注错误的偶然误差，通过分类模型对交叉验证集进行多次标注，并将标注结果和预设标准问进行对比，可以有效增强异常扩展问的识别准确性。

可选地，在获取分类模型的输出的所有的扩展问的M次标注结果之后，方法还包括：

当标注结果与预设标准问相同时，由标注结果与预设标准问所组成的标准问对的当次计分为第一预设分值；或，当标注结果与预设标准问不同时，标准问对的当次计分为第二预设分值；计算各个标准问对的第二累计得分；筛选出第二累计得分小于第二预设阈值的标准问对；将筛选出的标准问对作为易混淆标准问对，并输出。

例如，分类模型标注10次后，10个预设标准问（A1，A2...A10），10种标注结果（a1，a2...a10），就会有100种可能的标注问对，如：标准问对（A1，a1）。当扩展问的标注结果a1与预设标准问A1相同时，标准问对（A1，a1）的得分为1，不同则得分-1；分类模型标注10次后，标准问对（A1，a1）的得分越低，预设标准问A1与标注结果a1之间越容

易混淆。将易混淆标准问对输出，有利于开发人员修正标注结果，精确分类，降低混淆概率。

可选地，在步骤S105之后，方法还包括：

获取实质相同的多个扩展问的标注结果，实质相同的多个扩展问的语义相似度大于预设值；统计实质相同的多个扩展问的标注结果；将重复次数最多的标注结果作为实质相同的多个扩展问对应的正确预设标准问；将标注结果与正确预设标准问不同的扩展问确认为异常扩展问。

其中，计算语义相似度的方法有余弦相似度计算法、BM25算法等。

在一种实施方式中，计算语义相似度的方法包括：获取同一预设标准问关联的多个扩展问，将每个扩展问进行分词处理并提取关键词，其中，可以采用分词开源工具，例如ICTCLAS(Institute of Computing Technology, Chinese Lexical Analysis System，汉语词法分析系统)、SCWS(Simple Chinese Words Segmentation，简易中文分词系统)等，也可以直接使用自行开发的分词接口对每个子段文本进行分词；通过词向量表示模型获取每个关键词的向量表示，其中，词向量表示模型例如可以是Word2vec等神经网络模型。采用余弦相似度公式分别计算任意两个扩展问的相似度，将相似度大于预设阈值的多个扩展问确认为实质相同的多个扩展问，例如仅标点符号差异的扩展问。

可选地，在将标注结果与正确预设标准问不同的扩展问确认为异常扩展问之后，方法还包括：将异常扩展问所关联的预设标准问修改为正确预设标准问。

可选地，方法还包括：修正异常扩展问，并用修正后的异常扩展问替换原扩展问；用替换后的多个扩展问训练分类模型。

可以理解地，用修正后的异常扩展问替换原扩展问，使得分类模型的训练集中的样本精准度更高，从而进一步保障分类模型输出的标注更加精确。

本申请实施例提供了一种训练样本有效性检测装置，该训练样本有效性检测装置用于执行上述训练样本有效性检测方法，如图2所示，该装置包括：第一获取单元10、制作单元20、训练单元30、标注单元40、第二

获取单元 50、第一处理单元 60。

第一获取单元10，用于获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

制作单元20，用于将多个扩展问随机分为预设份数的样本集，并根据预设比例将预设份数的样本集分为训练集和交叉验证集；

训练单元30，用于利用训练集训练分类模型；

标注单元40，用于采用交叉验证方法，并利用分类模型对交叉验证集中的多个扩展问标注，直至所有的扩展问标注完成；

第二获取单元50，用于获取分类模型输出的所有的扩展问的标注结果；

第一处理单元60，用于根据标注结果得到异常扩展问，异常扩展问的标注结果与所关联的预设标准问不同。

在本方案中，利用交叉验证方法和分类模型来对扩展问进行标注，从而根据标注结果与扩展问关联的预设标准问进行匹配，得到异常扩展问，快速筛选出异常扩展问，提高了训练样本的有效性检测效率，从而避免一些质量低的扩展问来训练智能应答模型。

例如，预设标准问例如为“信用卡逾期怎么办？”，它的扩展问 1 为“我的信用卡逾期了，需要怎么处理”，扩展问 2 为“信用卡逾期可以不还吗”，扩展问 3 为“过了信用卡还款日，怎么办”。可以理解地，扩展问 1、扩展问 2、扩展问 3 所关联的预设标准问皆为“信用卡逾期怎么办？”

在本实施例中，训练集用于对分类模型进行训练样本集，为了使训练后的分类模型能够基于用户输入的问题识别出用户可能期望咨询的标准问。训练集中的扩展问为待分类的数据，其类别即为其关联的预设标准问。具体地，分类模型包括输入层、输出层及融合层。其中，输入层包括词语特征输入层和问题特征输入层。每个扩展问所关联的预设标准问作为问题特征，问题特征输入层用于获取输入的扩展问关联的预设标准问的特征，词语特征输入层用于获取扩展问分词处理后得到的词汇的特征。分类模型在经过训练后，能够根据学习到的词汇特性及问题特征来对一个新的扩展问进行分类，使得新的扩展问尽可能地分

类至一个准确的标准问下。

可选地，交叉验证方法为十折交叉验证方法或五折交叉验证方法中的任意一种。在一种实施方式中，采用五折交叉验证方法，具体地，将多个扩展问随机分成10份，每次取其中2份作为交叉验证集，其余8份作为训练集。训练时，先用其中的8份对初始化后的分类模型进行训练，然后对2份交叉验证集进行分类标注，以此重复训练及验证过程5次，每次选取的交叉验证集不同，直至所有的扩展问都被标注一遍。可以理解地，根据分类模型的分类结果对交叉验证集中的扩展问进行标注，例如分类模型将扩展问4“我的信用卡要逾期了，我该怎么处理”分类至预设标准问“信用卡逾期怎么办”下，那么扩展问4的关联标准问就是分类模型的分类结果。

在另一种实施方式中，将每个预设标准问所关联的多个扩展问随机分成10份，每次从每个预设标准问所关联的10份扩展问中抽取1份作为交叉验证集，其余9份作为训练集。例如：预设标准问有10个，分别为A1，A2...A10，每个预设标准问关联多个扩展问，因此将每个预设标准问所关联的多个扩展问随机分成10份，可以保证交叉验证时每个预设标准问都能得到有效的训练。

可选地，装置还包括重复单元、第二获取单元、计分单元、计算单元、筛选单元、输出单元。

重复单元，用于控制所述制作单元20、所述训练单元30及所述标注单元40重复操作M次，其中，M为大于1的自然数，每次重复操作前对分类模型进行初始化设置；第二获取单元，还用于获取分类模型的输出的所有的扩展问的M次标注结果；计分单元，用于当扩展问的标注结果与预设标准问相同时，扩展问的当次计分为第一预设分值；或，当扩展问的标注结果与预设标准问不同时，扩展问的当次计分为第二预设分值；计算单元，用于计算各个扩展问的第一累计得分；筛选单元，用于筛选出第一累计得分小于第一预设阈值的扩展问；输出单元，用于将筛选出的扩展问作为异常扩展问，并输出。

具体地，分类模型标注M次后，对M次的所有扩展问的标注结果进行统计，扩展问的得分越低，其关联的预设标准问是错误的可能性越高。例

如：当标注结果与预设标准问相同时，扩展问的得分为1；当标注结果与预设标准问不同时，扩展问的得分为-1，重复标注10次，其中6次得分为1，4次得分为-1，则扩展问的得分为10次标注结果的加和平均值，为0.2分。当第一预设阈值为0.1分时，则表示多个扩展问的分值小于0.1分的皆为异常扩展问。

可选地，所述分类模型可以采用多种机器学习算法，可以是朴素贝叶斯算法、最近邻法则、线性判别分析、支持向量机或决策树算法。在重复操作M次时，所述分类模型所采用的机器学习算法可以相同，也可以不同。采用不同的机器学习算法，能够减小标注错误的偶然误差，通过分类模型对交叉验证集进行多次标注并将标注结果和预设标准问进行对比，可以有效增强异常扩展问的识别准确性。

可选地，计分单元，还用于当标注结果与预设标准问相同时，由标注结果与预设标准问所组成的标准问对的当次计分为第一预设分值；或，当标注结果与预设标准问不同时，标准问对的当次计分为第二预设分值；计算单元，还用于计算各个标准问对的第二累计得分；筛选单元，还用于筛选出第二累计得分小于第二预设阈值的标准问对；输出单元，还用于将筛选出的标准问对作为易混淆标准问对，并输出。

例如，分类模型标注10次后，10个预设标准问（A1，A2...A10），10种标注结果（a1，a2...a10），就会有100种可能的标注问对，如：标准问对（A1，a1）。当扩展问的标注结果a1与预设标准问A1相同时，标准问对（A1，a1）的得分为1，不同则得分-1；分类模型标注10次后，标准问对（A1，a1）的得分越低，预设标准问A1与标注结果a1之间越容易混淆。将易混淆标准问对输出，有利于开发人员修正标注结果，精确分类，降低混淆概率。

可选地，装置还包括第三获取单元、统计单元、第二处理单元、确认单元。

第三获取单元，用于获取实质相同的多个扩展问的标注结果，实质相同的多个扩展问的语义相似度大于预设值；统计单元，用于统计实质相同的多个扩展问的标注结果；第二处理单元，用于将重复次数最多的标注结

果作为实质相同的多个扩展问对应的正确预设标准问；确认单元，用于将标注结果与正确预设标准问不同的扩展问确认为异常扩展问。

其中，计算语义相似度的方法有余弦相似度计算法、BM25算法等。

在一种实施方式中，计算语义相似度的方法包括：获取同一预设标准问关联的多个扩展问，将每个扩展问进行分词处理并提取关键词，其中，可以采用分词开源工具，例如ICTCLAS、SCWS等，也可以直接使用自行开发的分词接口对每个子段文本进行分词；通过词向量表示模型获取每个关键词的向量表示，其中，词向量表示模型例如可以是WORD2VEC等神经网络模型。采用余弦相似度公式分别计算任意两个扩展问的相似度，将相似度大于预设阈值的多个扩展问确认为实质相同的多个扩展问，例如仅标点符号差异的扩展问。

可选地，装置还包括修改单元，用于将异常扩展问所关联的预设标准问修改为正确预设标准问。

可选地，装置还包括修正单元和替换单元；

修正单元，用于修正异常扩展问，并用修正后的异常扩展问替换原扩展问；替换单元，用于用替换后的多个扩展问训练分类模型。

可以理解地，用修正后的异常扩展问替换原扩展问，使得分类模型的训练集中的样本精准度更高，从而进一步保障分类模型输出的标注更加精确。

本申请实施例提供了一种计算机非易失性存储介质，存储介质包括存储的程序，其中，在程序运行时控制存储介质所在设备执行以下步骤：

步骤S101，获取多个扩展问，其中，每个扩展问与对应的一个预设标准问关联；步骤S102，将多个扩展问随机分为预设份数的样本集，并根据预设比例将预设份数的样本集分为训练集和交叉验证集；步骤S103，利用训练集训练分类模型；步骤S104，采用交叉验证方法，并利用分类模型对交叉验证集中的多个扩展问标注，直至所有的扩展问标注完成；步骤S105，获取分类模型输出的所有的扩展问的标注结果；步骤S106，根据标注结果得到异常扩展问，异常扩展问的标注结果与所关联的预设标准问不同。

可选地，在程序运行时控制存储介质所在设备还执行以下步骤：重复操作M次的步骤S102至步骤S104，其中，M为大于1的自然数，每次重复操作前对分类模型进行初始化设置；获取分类模型的输出的所有的扩展问的M次标注结果；当扩展问的标注结果与预设标准问相同时，扩展问的当次计分为第一预设分值；或，当扩展问的标注结果与预设标准问不同时，扩展问的当次计分为第二预设分值；计算各个扩展问的第一累计得分；筛选出第一累计得分小于第一预设阈值的扩展问；将筛选出的扩展问作为异常扩展问，并输出。

可选地，在程序运行时控制存储介质所在设备在执行获取分类模型的输出的所有的扩展问的M次标注结果之后，还包括：当标注结果与预设标准问相同时，由标注结果与预设标准问所组成的标准问对的当次计分为第一预设分值；或，当标注结果与预设标准问不同时，标准问对的当次计分为第二预设分值；计算各个标准问对的第二累计得分；筛选出第二累计得分小于第二预设阈值的标准问对；将筛选出的标准问对作为易混淆标准问对，并输出。

可选地，在程序运行时控制存储介质所在设备在执行步骤S105之后，还包括：

获取实质相同的多个扩展问的标注结果，实质相同的多个扩展问的语义相似度大于预设值；统计实质相同的多个扩展问的标注结果；将重复次数最多的标注结果作为实质相同的多个扩展问对应的正确预设标准问；将标注结果与正确预设标准问不同的扩展问确认为异常扩展问。

可选地，在程序运行时控制存储介质所在设备在执行将标注结果与正确预设标准问不同的扩展问确认为异常扩展问之后，还包括：

将异常扩展问所关联的预设标准问修改为正确预设标准问。

图3是本申请实施例提供的一种计算机设备的示意图。如图3所示，该实施例的计算机设备100包括：处理器101、存储器102以及存储在存储器102中并可在处理器101上运行的计算机程序103，该计算机程序103被处理器101执行时实现实施例中的训练样本有效性检测方法，为避免重复，此处不一一赘述。或者，该计算机程序被处理器101执行时实

现实施例训练中训练样本有效性检测装置中各模型/单元的功能，为避免重复，此处不一一赘述。

计算机设备 100 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。计算机设备可包括，但不仅限于，处理器 101、存储器 102。本领域技术人员可以理解，图 3 仅仅是计算机设备 100 的示例，并不构成对计算机设备 100 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如计算机设备还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 101 可以是中央处理单元 (Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

存储器 102 可以是计算机设备 100 的内部存储单元，例如计算机设备 100 的硬盘或内存。存储器 102 也可以是计算机设备 100 的外部存储设备，例如计算机设备 100 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，存储器 102 还可以既包括计算机设备 100 的内部存储单元也包括外部存储设备。存储器 102 用于存储计算机程序以及计算机设备所需的其他程序和数据。存储器 102 还可以用于暂时地存储已经输出或者将要输出的数据。

需要说明的是，本申请实施例中所涉及的终端可以包括但不限于个人计算机 (Personal Computer, PC)、个人数字助理 (Personal Digital Assistant, PDA)、无线手持设备、平板电脑 (Tablet Computer)、手机、MP3 播放器、MP4 播放器等。

可以理解的是，所述应用可以是安装在终端上的应用程序 (native App)，或者还可以是终端上的浏览器的一个网页程序 (web App)，本申

请实施例对此不进行限定。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如，多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机装置（可以是个人计算机，服务器，或者网络装置等）或处理器（Processor）执行本申请各个实施例所述方法的部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述仅为本申请的较佳实施例而已，并不用以限制本申请，凡在本申请的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本申请保护的范围之内。

权 利 要 求 书

1、一种训练样本有效性检测方法，其特征在于，所述方法包括：

步骤 S101，获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

步骤 S102，将多个所述扩展问随机分为预设份数的样本集，并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集；

步骤 S103，利用所述训练集训练分类模型；

步骤 S104，采用交叉验证方法，并利用所述分类模型对所述交叉验证集中的多个扩展问标注，直至所有的所述扩展问标注完成；

步骤 S105，获取所述分类模型输出的所有的所述扩展问的标注结果；

步骤 S106，根据所述标注结果得到异常扩展问，所述异常扩展问的标注结果与所关联的预设标准问不同。

2、根据权利要求 1 所述的方法，其特征在于，在所述步骤 S105 之后，所述方法还包括：

重复操作 M 次的所述步骤 S102 至所述步骤 S104，其中，M 为大于 1 的自然数，每次重复操作前对所述分类模型进行初始化设置；

获取所述分类模型的输出的所有的所述扩展问的 M 次标注结果；

当所述扩展问的标注结果与所述预设标准问相同时，所述扩展问的当次计分为第一预设分值；或，当所述扩展问的标注结果与所述预设标准问不同时，所述扩展问的当次计分为第二预设分值；

计算各个所述扩展问的第一累计得分；

筛选出所述第一累计得分小于第一预设阈值的所述扩展问；

将筛选出的扩展问作为所述异常扩展问，并输出。

3、根据权利要求 2 所述的方法，其特征在于，在所述获取所述分类模型的输出的所有的所述扩展问的 M 次标注结果之后，所述方法还包括：

当所述标注结果与所述预设标准问相同时，由所述标注结果与所述预设标准问所组成的标准问对的当次计分为所述第一预设分值；或，当所述标注结果与所述预设标准问不同时，所述标准问对的当次计分为所述第二预设分值；

计算各个所述标准问对的第二累计得分；

筛选出所述第二累计得分小于第二预设阈值的所述标准问对；

将筛选出的标准问对作为易混淆标准问对，并输出。

4、根据权利要求 1 所述的方法，其特征在于，在所述步骤 S105 之后，所述方法还包括：

获取实质相同的多个扩展问的标注结果，所述实质相同的多个扩展问的语义相似度大于预设值；

统计所述实质相同的多个扩展问的标注结果；

将重复次数最多的标注结果作为所述实质相同的多个扩展问对应的

正确预设标准问；

将所述标注结果与所述正确预设标准问不同的扩展问确认为所述异常扩展问。

5、根据权利要求4所述的方法，其特征在于，在所述将所述标注结果与所述正确预设标准问不同的扩展问确认为所述异常扩展问之后，所述方法还包括：

将所述异常扩展问所关联的预设标准问修改为所述正确预设标准问。

6、根据权利要求1~5中任意一项所述的方法，其特征在于，所述方法还包括：

修正所述异常扩展问，并用修正后的所述异常扩展问替换原扩展问；用替换后的多个所述扩展问训练所述分类模型。

7、根据权利要求1所述的方法，其特征在于，所述交叉验证方法为十折交叉验证方法或五折交叉验证方法中的任意一种。

8、一种训练样本有效性检测装置，其特征在于，所述装置包括：

第一获取单元，用于获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

制作单元，用于将多个所述扩展问随机分为预设份数的样本集，并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集；

训练单元，用于利用所述训练集训练分类模型；

标注单元，用于采用交叉验证方法，并利用所述分类模型对所述交叉验证集中的多个扩展问标注，直至所有的所述扩展问标注完成；

第二获取单元，用于获取所述分类模型输出的所有的所述扩展问的标注结果；

第一处理单元，用于根据所述标注结果得到异常扩展问，所述异常扩展问的标注结果与所关联的预设标准问不同。

9、根据权利要求8所述的装置，其特征在于，所述装置还包括：

重复单元，用于控制所述制作单元、所述训练单元及所述标注单元重复操作M次，其中，M为大于1的自然数，每次重复操作前对所述分类模型进行初始化设置；

所述第二获取单元，还用于获取所述分类模型的输出的所有的所述扩展问的M次标注结果；

计分单元，用于当所述扩展问的标注结果与所述预设标准问相同时，所述扩展问的当次计分为第一预设分值；或，当所述扩展问的标注结果与所述预设标准问不同时，所述扩展问的当次计分为第二预设分值；

计算单元，用于计算各个所述扩展问的第一累计得分；

筛选单元，用于筛选出所述第一累计得分小于第一预设阈值的扩展问；

输出单元，用于将筛选出的扩展问作为所述异常扩展问，并输出。

10、根据权利要求9所述的装置，其特征在于，

所述计分单元，还用于当所述标注结果与所述预设标准问相同时，由

所述标注结果与所述预设标准问所组成的标准问对的当次计分为所述第一预设分值；或，当所述标注结果与所述预设标准问不同时，所述标准问对的当次计分为所述第二预设分值；

所述计算单元，还用于计算各个所述标准问对的第二累计得分；

所述筛选单元，还用于筛选出所述第二累计得分小于第二预设阈值的所述标准问对；

所述输出单元，还用于将筛选出的标准问对作为易混淆标准问对，并输出。

11、根据权利要求 8 所述的装置，其特征在于，所述装置还包括：

第三获取单元，用于获取实质相同的多个扩展问的标注结果，所述实质相同的多个扩展问的语义相似度大于预设值；

统计单元，用于统计所述实质相同的多个扩展问的标注结果；

第二处理单元，用于将重复次数最多的标注结果作为所述实质相同的多个扩展问对应的正确预设标准问；

确认单元，用于将所述标注结果与所述正确预设标准问不同的扩展问确认为所述异常扩展问。

12、根据权利要求 8~11 任一项所述的装置，其特征在于，所述装置还包括：

修正单元，用于修正异常扩展问，并用修正后的异常扩展问替换原扩展问；

替换单元，用于用替换后的多个扩展问训练分类模型。

13、一种计算机非易失性存储介质，所述存储介质包括存储的程序，其特征在于，在所述程序运行时控制所述存储介质所在设备执行以下步骤：

步骤 S101，获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

步骤 S102，将多个所述扩展问随机分为预设份数的样本集，并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集；

步骤 S103，利用所述训练集训练分类模型；

步骤 S104，采用交叉验证方法，并利用所述分类模型对所述交叉验证集中的多个扩展问标注，直至所有的所述扩展问标注完成；

步骤 S105，获取所述分类模型输出的所有的所述扩展问的标注结果；

步骤 S106，根据所述标注结果得到异常扩展问，所述异常扩展问的标注结果与所关联的预设标准问不同。

14、根据权利要求 13 所述的计算机非易失性存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备执行以下步骤：

重复操作 M 次的所述步骤 S102 至所述步骤 S104，其中，M 为大于 1 的自然数，每次重复操作前对所述分类模型进行初始化设置；

获取所述分类模型的输出的所有的所述扩展问的 M 次标注结果；

当所述扩展问的标注结果与所述预设标准问相同时，所述扩展问的当

次计分为第一预设分值；或，当所述扩展问的标注结果与所述预设标准问不同时，所述扩展问的当次计分为第二预设分值；

计算各个所述扩展问的第一累计得分；

筛选出所述第一累计得分小于第一预设阈值的所述扩展问；

将筛选出的扩展问作为所述异常扩展问，并输出。

15、根据权利要求 14 所述的计算机非易失性存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备在执行所述获取所述分类模型的输出的所有的所述扩展问的 M 次标注结果之后，还包括：

当所述标注结果与所述预设标准问相同时，由所述标注结果与所述预设标准问所组成的标准问对的当次计分为所述第一预设分值；或，当所述标注结果与所述预设标准问不同时，所述标准问对的当次计分为所述第二预设分值；

计算各个所述标准问对的第二累计得分；

筛选出所述第二累计得分小于第二预设阈值的所述标准问对；

将筛选出的标准问对作为易混淆标准问对，并输出。

16、根据权利要求 13 所述的计算机非易失性存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备在执行步骤 S105 之后，还包括：

获取实质相同的多个扩展问的标注结果，所述实质相同的多个扩展问的语义相似度大于预设值；

统计所述实质相同的多个扩展问的标注结果；

将重复次数最多的标注结果作为所述实质相同的多个扩展问对应的正确预设标准问；

将所述标注结果与所述正确预设标准问不同的扩展问确认为所述异常扩展问。

17、一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现以下步骤：

步骤 S101，获取多个扩展问，其中，每个所述扩展问与对应的一个预设标准问关联；

步骤 S102，将多个所述扩展问随机分为预设份数的样本集，并根据预设比例将所述预设份数的样本集分为训练集和交叉验证集；

步骤 S103，利用所述训练集训练分类模型；

步骤 S104，采用交叉验证方法，并利用所述分类模型对所述交叉验证集中的多个扩展问标注，直至所有的所述扩展问标注完成；

步骤 S105，获取所述分类模型输出的所有的所述扩展问的标注结果；

步骤 S106，根据所述标注结果得到异常扩展问，所述异常扩展问的标注结果与所关联的预设标准问不同。

18、根据权利要求 17 所述的计算机设备，其特征在于，所述处理器

执行所述计算机程序时还实现以下步骤：

重复操作 M 次的步骤 S102 至步骤 S104，其中，M 为大于 1 的自然数，每次重复操作前对分类模型进行初始化设置；

获取分类模型的输出的所有的扩展问的 M 次标注结果；

当扩展问的标注结果与预设标准问相同时，扩展问的当次计分为第一预设分值；或，当扩展问的标注结果与预设标准问不同时，扩展问的当次计分为第二预设分值；

计算各个所述扩展问的第一累计得分；

筛选出所述第二累计得分小于第二预设阈值的所述标准问对；

将筛选出的标准问对作为易混淆标准问对，并输出。

19、根据权利要求 18 所述的计算机设备，其特征在于，所述处理器执行所述计算机程序时在实现所述获取所述分类模型的输出的所有的所述扩展问的 M 次标注结果之后，还实现以下步骤：

当标注结果与预设标准问相同时，由标注结果与预设标准问所组成的标准问对的当次计分为第一预设分值；或，当标注结果与预设标准问不同时，标准问对的当次计分为第二预设分值；计算各个标准问对的第二累计得分；筛选出第二累计得分小于第二预设阈值的标准问对；将筛选出的标准问对作为易混淆标准问对，并输出。

20、根据权利要求 17 所述的计算机设备，其特征在于，所述处理器执行所述计算机程序时在实现步骤 S105 之后，还实现以下步骤：

获取实质相同的多个扩展问的标注结果，所述实质相同的多个扩展问的语义相似度大于预设值；

统计所述实质相同的多个扩展问的标注结果；

将重复次数最多的标注结果作为所述实质相同的多个扩展问对应的正确预设标准问；

将所述标注结果与所述正确预设标准问不同的扩展问确认为所述异常扩展问。

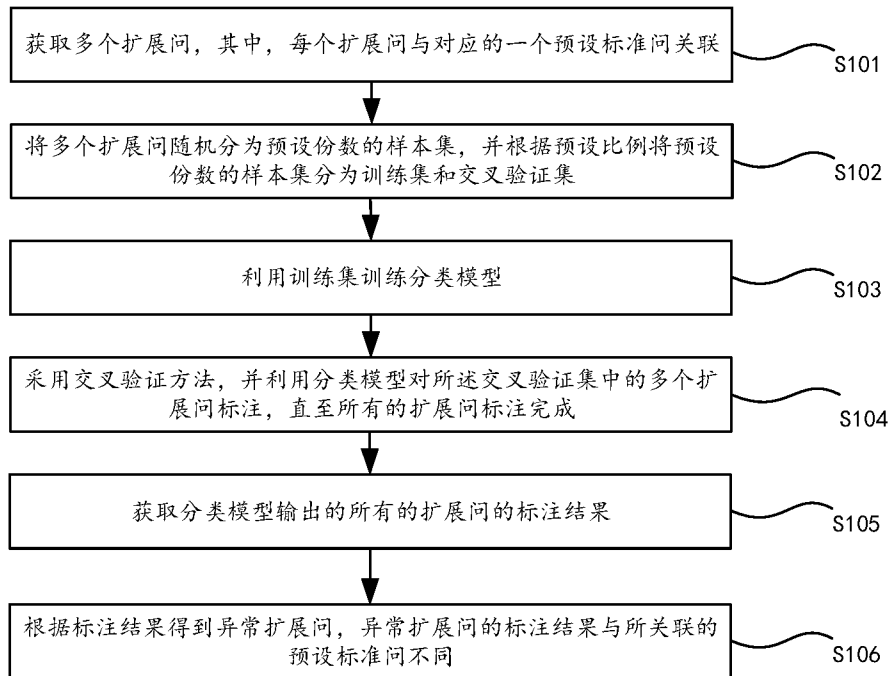


图 1

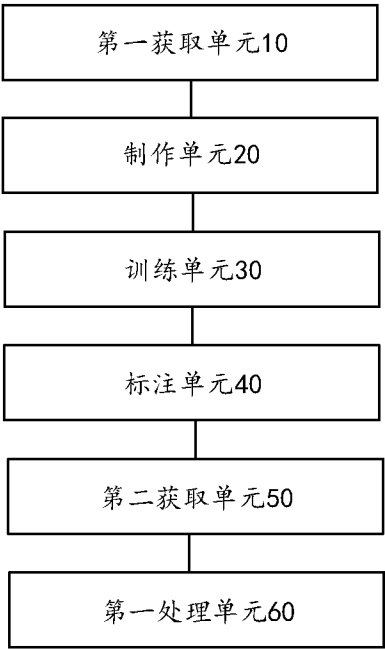


图 2

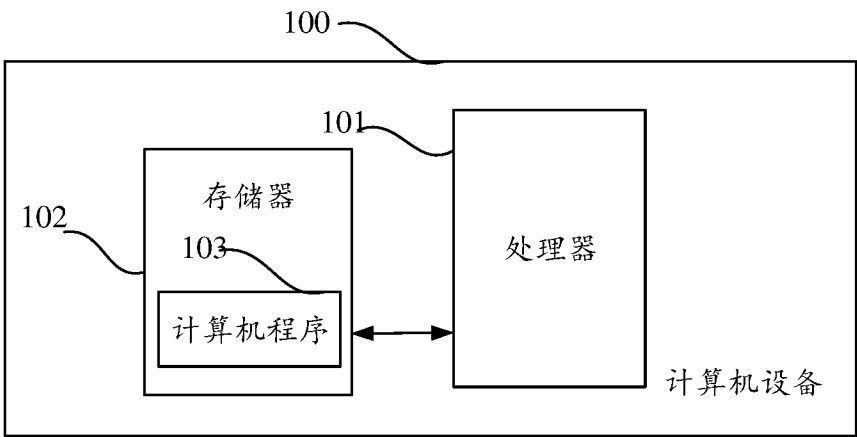


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/117090

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/332(2019.01)i; G06F 16/35(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)

G06F; G06N; G06K

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)

CNPAT, EPODOC, WPI, CNKI, IEEE: 检测, 训练样本, 样本, 扩展, 问题, 问句, 提问, 标注, 验证, 修正, 分类, 模型, 交叉, 预设, detection, training, sample, extension, extend, expand, question, sentence, label, verification, correction, classification, model, cross, presupposition

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109902157 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 18 June 2019 (2019-06-18) claims 1-10	1-20
X	CN 108734296 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.) 02 November 2018 (2018-11-02) description, paragraphs [0002]-[0023]	1-20
A	CN 108509969 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 07 September 2018 (2018-09-07) entire document	1-20
A	CN 108009589 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 08 May 2018 (2018-05-08) entire document	1-20



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

“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 January 2020

Date of mailing of the international search report

21 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/117090

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109902157	A	18 June 2019	None	
CN	108734296	A	02 November 2018	None	
CN	108509969	A	07 September 2018	None	
CN	108009589	A	08 May 2018	None	

国际检索报告

国际申请号

PCT/CN2019/117090

A. 主题的分类 G06F 16/332(2019.01)i; G06F 16/35(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) G06F; G06N; G06K 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)) CNPAT, EPDOC, WPI, CNKI, IEEE:检测, 训练样本, 样本, 扩展, 问题, 问句, 提问, 标注, 验证, 修正, 分类, 模型, 交叉, 预设, detection, training, sample, extension, extend, expand, question, sentence, label, verification, correction, classification, model, cross, presupposition																	
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 109902157 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 权利要求1-10</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 108734296 A (北京京东尚科信息技术有限公司 等) 2018年 11月 2日 (2018 - 11 - 02) 说明书第[0002]-[0023]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108509969 A (腾讯科技深圳有限公司) 2018年 9月 7日 (2018 - 09 - 07) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108009589 A (腾讯科技深圳有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文</td> <td>1-20</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 109902157 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 权利要求1-10	1-20	X	CN 108734296 A (北京京东尚科信息技术有限公司 等) 2018年 11月 2日 (2018 - 11 - 02) 说明书第[0002]-[0023]段	1-20	A	CN 108509969 A (腾讯科技深圳有限公司) 2018年 9月 7日 (2018 - 09 - 07) 全文	1-20	A	CN 108009589 A (腾讯科技深圳有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
PX	CN 109902157 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 权利要求1-10	1-20															
X	CN 108734296 A (北京京东尚科信息技术有限公司 等) 2018年 11月 2日 (2018 - 11 - 02) 说明书第[0002]-[0023]段	1-20															
A	CN 108509969 A (腾讯科技深圳有限公司) 2018年 9月 7日 (2018 - 09 - 07) 全文	1-20															
A	CN 108009589 A (腾讯科技深圳有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文	1-20															
☐ 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 “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 2020年 1月 14日		国际检索报告邮寄日期 2020年 2月 21日															
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 吴媛媛 电话号码 86-(10)-53961351															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/117090

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	109902157	A	2019年 6月 18日	无	
CN	108734296	A	2018年 11月 2日	无	
CN	108509969	A	2018年 9月 7日	无	
CN	108009589	A	2018年 5月 8日	无	