

(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 8 月 27 日 (27.08.2020)



(10) 国际公布号
WO 2020/168843 A1

(51) 国际专利分类号:
G06N 20/00 (2019.01)

(21) 国际申请号: PCT/CN2020/070290

(22) 国际申请日: 2020 年 1 月 3 日 (03.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910133409.3 2019年2月22日 (22.02.2019) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (**ALIBABA GROUP HOLDING LIMITED**) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 林建滨 (**LIN, Jianbin**); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (**BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

(54) **Title:** MODEL TRAINING METHOD AND APPARATUS BASED ON DISTURBANCE SAMPLES

(54) 发明名称: 一种基于扰动样本的模型训练方法和装置

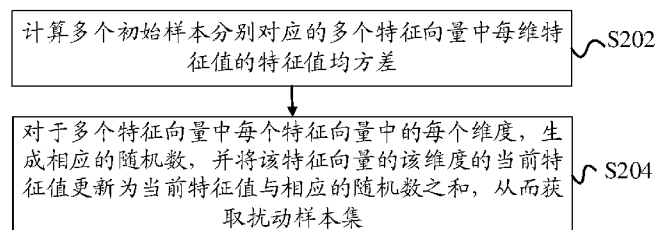


图 2

- S202 Calculate a mean square error of feature values of each dimension in a plurality of feature vectors respectively corresponding to a plurality of initial samples
- S204 For each dimension in each of the plurality of feature vectors, generate a corresponding random number, and update the current feature value of the dimension of the feature vector to be the sum of the current feature value and the corresponding random number, and thereby acquire a disturbance sample set

(57) **Abstract:** Provided are a method and apparatus for acquiring a disturbance sample set and training a model on the basis of the disturbance sample set. The method for acquiring a disturbance sample set comprises: calculating a mean square error of feature values of each dimension in a plurality of feature vectors respectively corresponding to a plurality of initial samples (S202); and for each dimension in each of the plurality of feature vectors, generating a corresponding random number, and updating the current feature value of the dimension of the feature vector to be the sum of the current feature value and the corresponding random number so as to generate a plurality of disturbance samples respectively corresponding to the plurality of feature vectors, and thereby acquiring a disturbance sample set (S204), wherein a value range of each random number is determined based on the product of a predetermined first parameter and a mean square error of feature values of a dimension corresponding to the random number.

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))。

(57) 摘要：一种获取扰动样本集及基于扰动样本集训练模型的方法和装置，所述获取扰动样本集的方法包括：计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差(S202)；以及对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集(S204)，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

一种基于扰动样本的模型训练方法和装置

技术领域

[01] 本说明书实施例涉及机器学习领域，更具体地，涉及基于初始样本集获取扰动样本集的方法和装置、获取模型训练样本集的方法和装置、获取模型训练样本集和测试样本集的方法和装置、以及基于扰动样本的模型训练方法和装置。

背景技术

[02] 机器模型部署到实际环境中会受到各种各样的挑战，其中很重要的一个挑战就是模型的稳定性。以语音识别模型为例，机器学习模型在训练时候用到的数据往往是经过妥善处理、降噪的，然而在实际环境中，模型要面对的情况却十分复杂，比如处于嘈杂的环境、话筒的回声等都会导致模型要处理的数据带有噪声，和实际的训练数据不一致，从而导致模型的精度发生较大的变化。因此，提高机器学习模型的鲁棒性对于机器学习模型的实际应用有着重大的意义。目前机器学习算法普遍采用 L1 正则和 L2 正则来增强模型的鲁棒性，这两个方法都是通过限制模型参数的搜索空间（参数的绝对值）达到鲁棒性的效果。

[03] 因此，需要一种更有效的增强模型鲁棒性的模型训练方法。

发明内容

[04] 本说明书实施例旨在提供一种更有效的模型训练方法，以解决现有技术中的不足。

[05] 为实现上述目的，本说明书一个方面提供一种基于初始样本集获取扰动样本集的方法，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述方法包括：

[06] 计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

[07] 对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积

确定。

[08] 在一个实施例中，所述随机数为高斯分布随机数，所述高斯分布随机数的均方差为所述第一参数与该随机数对应的维度的特征值均方差的乘积。

5 [09] 在一个实施例中，所述随机数为平均随机数，其中，所述平均随机数的取值范围在正负第一数值之间，其中，所述第一数值为所述第一参数与该随机数对应的维度的特征值均方差的乘积。

[10] 本说明书另一方面提供一种基于初始样本集获取模型训练样本集的方法，其中，所述初始样本集中包括多个初始样本，所述方法包括：

10 [11] 通过上述方法，获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；以及

[12] 通过将所述多个初始样本与所述多个扰动样本合并，获取训练样本集。

[13] 本说明书另一方面提供一种基于初始样本集获取模型训练样本集和测试样本集的方法，其中，所述初始样本集中包括多个初始样本，所述方法包括：

15 [14] 通过上述获取扰动样本集的方法，获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；

[15] 通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并，获取训练样本集；以及

[16] 通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集。

20 [17] 在一个实施例中，所述部分初始样本占所述多个初始样本的比例与所述部分扰动样本占所述多个扰动样本的比例相同。

[18] 在一个实施例中，通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集包括，通过将所述多个初始样本中剩余的初始样本与所述多个扰动样本中剩余的扰动样本合并，获取测试样
25 本集。

[19] 在一个实施例中，所述部分初始样本与所述部分扰动样本分别对应。

[20] 本说明书实施例提供一种模型训练方法，包括：

[21] 获取初始样本集，其中，所述初始样本集中包括多个初始样本；

[22] 通过上述获取训练样本集和测试样本集的方法基于所述初始样本集获取多个训练样本集、及与所述多个训练样本集分别对应的多个测试样本集，其中，所述多个训练样本集与多个取值不同的第一参数分别对应；

5 [23] 使用所述多个训练样本集分别训练当前模型，以分别获取多个更新模型；

[24] 使用所述多个测试样本集分别评估相应的更新模型，其中，所述测试样本集与相应的更新模型对应于相同的训练样本集；以及

[25] 基于评估结果，在所述多个更新模型中确定所述当前模型的更新模型。

10 [26] 在一个实施例中，所述模型为以下任一类模型：监督学习模型、无监督学习模型、和强化学习模型。

[27] 本说明书另一方面提供一种基于初始样本集获取扰动样本集的装置，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述装置包括：

[28] 计算单元，配置为，计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

15 [29] 生成单元，配置为，对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

20 [30] 本说明书另一方面提供一种基于初始样本集获取模型训练样本集的装置，其中，所述初始样本集中包括多个初始样本，所述装置包括：

[31] 获取单元，配置为，通过上述装置，获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；以及

25 [32] 合并单元，配置为，通过将所述多个初始样本与所述多个扰动样本合并，获取训练样本集。

[33] 本说明书另一方面提供一种基于初始样本集获取模型训练样本集和测试样本集的装置，其中，所述初始样本集中包括多个初始样本，所述装置包括：

[34] 获取单元，配置为，通过上述获取扰动样本集的装置，获取扰动样本集，所述扰

动样本集包括与所述多个初始样本分别对应的多个扰动样本；

[35] 第一合并单元，配置为，通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并，获取训练样本集；以及

5 [36] 第二合并单元，配置为，通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集。

[37] 在一个实施例中，所述第二合并单元还配置为，通过将所述多个初始样本中剩余的初始样本与所述多个扰动样本中剩余的扰动样本合并，获取测试样本集。

[38] 本说明书另一方面提供一种模型训练装置，包括：

10 [39] 第一获取单元，配置为，获取初始样本集，其中，所述初始样本集中包括多个初始样本；

[40] 第二获取单元，配置为，通过上述获取训练样本集和测试样本集的装置基于所述初始样本集获取多个训练样本集、及与所述多个训练样本集分别对应的多个测试样本集，其中，所述多个训练样本集与多个取值不同的第一参数分别对应；

15 [41] 训练单元，配置为，使用所述多个训练样本集分别训练当前模型，以分别获取多个更新模型；

[42] 评估单元，配置为，使用所述多个测试样本集分别评估相应的更新模型，其中，所述测试样本集与相应的更新模型对应于相同的训练样本集；以及

[43] 确定单元，配置为，基于评估结果，在所述多个更新模型中确定所述当前模型的更新模型。

20 [44] 本说明书另一方面提供一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行上述任一项方法。

[45] 本说明书另一方面提供一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现上述任一项方法。

25 [46] 本说明书实施例通过对模型的训练数据进行扰动，模拟真实环境中的数据噪声，从而增加模型的鲁棒性，并且通过使用扰动数据训练并评估模型，并基于评估结果确定模型预定参数，从而定量地提高了模型对异常数据的有效性，另外，本说明书实施例中未对机器学习模型的参数进行限制，从而不会限制模型的学习能力。

附图说明

[47] 通过结合附图描述本说明书实施例，可以使得本说明书实施例更加清楚：

[48] 图 1 示出根据本说明书实施例的模型训练系统 100 的示意图；

[49] 图 2 示出根据本说明书实施例的一种基于初始样本集获取扰动样本集的方法；

5 [50] 图 3 示意示出对多个特征向量中的一维特征值的均方差的计算；

[51] 图 4 示出与图 3 中 n 个特征向量分别对应的 n 个扰动特征向量；

[52] 图 5 示出根据本说明书实施例的基于初始样本集获取模型训练样本集的方法流程图；

10 [53] 图 6 示出了根据本说明书实施例的一种基于初始样本集获取模型训练样本集和测试样本集的方法流程图；

[54] 图 7 示出根据本说明书实施例的一种模型训练方法流程图；

[55] 图 8 示出根据本说明书实施例的一种基于初始样本集获取扰动样本集的装置 800；

[56] 图 9 示出根据本说明书实施例的一种基于初始样本集获取模型训练样本集的装置 900；

15 [57] 图 10 示出根据本说明书实施例的一种基于初始样本集获取模型训练样本集和测试样本集的装置 1000；

[58] 图 11 示出根据本说明书实施例的一种模型训练装置 1100。

具体实施方式

[59] 下面将结合附图描述本说明书实施例。

20 [60] 图 1 示出根据本说明书实施例的模型训练系统 100 的示意图。如图 1 所示，系统 100 包括，数据处理模块 11、训练模块 12 和评估模块 13。在数据处理模块 11 中，通过对数据集 A 中的每个样本中的特征向量的每个维度值施加扰动，从而获取数据集 B。在训练模块 12，从数据集 A 中获取至少部分数据（例如数据集 A 中 70% 的数据），从数据集 B 获取至少部分数据（例如数据集 B 中 70% 的数据），从而通过将其合并而获取
25 训练数据集，并使用该训练数据集训练机器学习模型。该机器学习模型可以为任意模型，其例如为上文中的语音识别模型。该语音识别模型例如为监督学习模型或者为强化学习

模型等，可以理解，所述任意模型还可以为无监督学习模型。在评估模块 13，从数据集 A 中获取剩余数据（例如数据集 A 中的 30% 数据），从数据集 B 中获取剩余数据（例如数据集 B 中的 30% 数据），从而通过将其合并而获取测试数据集，并使用该测试数据集评估该训练的模型。

5 [61] 下面详细描述上述各个处理过程。

[62] 图 2 示出根据本说明书实施例的一种基于初始样本集获取扰动样本集的方法，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述方法包括：

[63] 在步骤 S202，计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

10 [64] 在步骤 S204，对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

15 [65] 首先，在步骤 S202，计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差。

[66] 所述初始样本集例如为图 1 中所示的数据集 A，在该数据集 A 中包括例如 n 个初始样本，每个样本中都包括各自的特征向量，所述特征向量例如为 m 维特征向量，其每个维度值对应于一个特征值。另外，在将要训练的模型例如为监督模型的情况中，所述

20 每个样本中还包括相应的标签值。

[67] 所述均方差即为标准差，其为方差 σ^2 的平方根，可用 σ 表示。在对多个样本计算方差时，通常用下面的公式（1）计算：

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2 \quad (1)$$

其中，n 为样本总数， μ 为 n 个 x_i 的均值。

25 [68] 图 3 示意示出对多个特征向量中的一维特征值的均方差的计算。在图 3 中，假设有 n 个特征向量，每个特征向量都包括 m 个特征的特征值，可将其中第 j 个特征向量（即图中向量 j）的第 i 维特征（即图中的特征 i）的特征值表示为 x_{ij} ，其中 $i \in [1, m]$ ， $j \in [1, n]$ 。从而，n 个特征向量各自的第 i 维的特征值的标准差 σ_i 可通过如下公式（2）计算：

$$\sigma_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \mu_i)^2} \quad (2)$$

其中, μ_i 通过如下公式 (3) 计算

$$\mu_i = \frac{\sum_{j=1}^n x_{ij}}{n} \quad (3)$$

具体是, 例如, 对于图 3 中虚线框中的各个向量的第 2 维特征 (即特征 2) 的特征值

5 $x_{21}, x_{22}, \dots, x_{2n}$, 可基于该 n 个数值, 通过公式 (3) 计算其均值 $\mu_2 = \frac{\sum_{j=1}^n x_{2j}}{n}$, 通过公式

$$(2) \text{ 计算其方差 } \sigma_2 = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{2j} - \mu_2)^2}.$$

[69] 在步骤 S204, 对于所述多个特征向量中每个特征向量中的每个维度, 生成相应的随机数, 并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和, 以生成与所述多个特征向量分别对应的多个扰动样本, 从而获取扰动样本集, 其中, 各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

[70] 仍然参考图 3, 假设对于图中的向量 1, 对于该特征向量的每个维度 $1, 2, \dots, m$, 分别生成相应的随机数 $a_{11}, a_{21}, \dots, a_{m1}$, 并将该特征向量每个维度的当前特征值 x_{i1} 更新为 $x_{i1} + a_{i1}$, 其中 $i \in [1, m]$, 从而可获取与该特征向量 1 对应的扰动特征向量 (如图 4 中虚线框内所示)。对于每个特征向量, 都可以类似地施加扰动, 从而获取与其对应的扰动特征向量。图 4 示出与图 3 中 n 个特征向量分别对应的 n 个扰动特征向量。通过该扰动特征向量训练模型, 即对训练样本增加了噪声, 将提高模型对噪声的抵抗能力, 增强模型的稳定性。

[71] 在一个实施例, 对于图 4 中包括的每个随机数 a_{ij} , 其中 $i \in [1, m]$, $j \in [1, n]$, 可通过高斯随机变量 $A = \text{norm}(0, \lambda \sigma_i)$ 获取, 也就是说该随机变量均值为 0, 均方差为 $\lambda \sigma_i$, 这里 σ_i 为通过上述公式 (2) 计算的维度 i (特征 i) 的特征值均方差, λ 为预定参数, 其可取值在 0.0001 到 0.1 之间。例如结合图 3 和图 4, 图 4 中与特征 2 维度对应的各个 a_{2j} 都通过 $A = \text{norm}(0, \lambda \sigma_2)$ 生成, 其中, σ_2 即为图 3 中所示 σ_2 。如本领域技术人员熟知的, 根据高斯变量的概率密度图形, a_{ij} 的 99% 的可能取值将落在 $[-3\lambda \sigma_i, 3\lambda \sigma_i]$ 的区间中, 也就是说, 通过该预定参数 λ 与维度 i 的特征值均方差 σ_i 的乘积 $\lambda \sigma_i$, 限定了维度 i 的各个随机数 a_{ij} 的取值范围, 其中 j 的取值范围为 1 到 n 。

[72] 在一个实施例, 图 4 中的各个随机数 a_{ij} 可通过平均随机变量 B 获取, 该随机变

量 B 例如可为 $[-\lambda\sigma_i, \lambda\sigma_i]$ 范围内的平均随机变量。也就是说，通过所述乘积 $\lambda\sigma_i$ 限定各个随机数 a_{ij} 的取值范围。所述平均随机变量 B 的取值范围不限于设定为 $[-\lambda\sigma_i, \lambda\sigma_i]$ ，例如，也可以为 $[-3\lambda\sigma_i, 3\lambda\sigma_i]$ 等等。另外，所述各个随机数 a_{ij} 不限于为通过上述高斯随机变量或平均随机变量获取，而可以通过其它任意随机变量获取，如泊松随机变量等等，只要其通过 $\lambda\sigma_i$ 限制其取值范围即可。

[73] 其中， λ 用于平衡加入的噪声和模型性能， λ 的值越小， a_{ij} 的取值范围越小，也即施加的扰动越小，当施加扰动太小时，对特征向量的影响太小，不能起到提高模型性能的作用，当施加的扰动过大时，又影响了模型的预测准确性。因此， λ 的取值对于模型性能的提高较为重要。在一个实施例中，可通过模型应用的具体环境确定 λ 的值，例如，对于语音识别模型，在其应用的环境比较嘈杂，噪音较大的情况中，可将 λ 值设置为较大，在其应用的环境比较安静，噪音偏小的情况中，可将 λ 值设置为较小。在一个实施例中，如下文将要详细描述，通过对训练好的模型进行评估来确定 λ 的取值，即，选择评估值较好的 λ 作为最终的模型训练的 λ 。在一个实施例中，当模型在使用之前批次的训练样本集训练中已经确定好 λ 之后，在后续的反复训练中可重复使用该 λ 值。

[74] 在通过图 2 所示方法获取初始样本集的各个初始样本分别对应的扰动样本之后，可获取扰动样本集，从而可基于初始样本集和扰动样本集获取模型的训练样本集和测试样本集。

[75] 图 5 示出根据本说明书实施例的基于初始样本集获取模型训练样本集的方法流程图，其中，所述初始样本集中包括多个初始样本，所述方法包括：

[76] 在步骤 S502，通过图 2 所示方法，基于所述初始样本集获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；以及

[77] 在步骤 S504，通过将所述多个初始样本与所述多个扰动样本合并，获取训练样本集。

[78] 首先，在步骤 S502，通过图 2 所示方法，生成与所述多个初始样本分别对应的多个扰动样本，其中，所述多个扰动样本对应于取值相同的第一参数。所述多个扰动样本例如为如图 4 所示的多个扰动样本，如上文所述，在该多个扰动样本中的每个随机数 a_{ij} 可通过高斯随机变量 $A = \text{norm}(0, \lambda\sigma_i)$ 获取，也就是说，该批扰动样本对应于相同的 λ 值。

[79] 在步骤 S504，通过将所述多个初始样本与所述多个扰动样本合并，获取训练样本

集。在不需要对训练模型进行评估的情况中，可将初始样本集中的全部样本与扰动样本集中的全部样本合并，从而获取训练样本集。通过在训练样本集中同时包括初始样本集和扰动样本集，丰富了模型的训练样本，使得模型能够适应于不同的实际环境。

[80] 图 6 示出了根据本说明书实施例的一种基于初始样本集获取模型训练样本集和测试样本集的方法流程图，其中，所述初始样本集中包括多个初始样本，所述方法包括：

[81] 在步骤 S602，通过图 2 所示方法，基于所述初始样本集获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；

[82] 在步骤 S604，通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并，获取训练样本集；以及

[83] 在步骤 S606，通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集。

[84] 在该方法中，在获取扰动样本集之后，可通过将部分初始样本与部分扰动样本合并在一起，从而获取训练样本集。在一个实施例中，例如，可合并初始样本集中的 70% 的初始样本和扰动样本集中的 70% 的扰动样本，从而获取训练样本集。之后，将所述初始样本集和扰动样本集中的剩余的例如 30% 的初始样本和 30% 的扰动样本合并，从而获取与该训练样本集对应的测试样本集。其中，所述训练样本集中的所述 70% 初始样本和所述 70% 扰动样本可以是分别相互对应的，也可以是不对应的。

[85] 在一个实施例中，所述部分初始样本占所述多个初始样本的比例与所述部分扰动样本占所述多个扰动样本的比例也可以不同，例如，在模型实际应用环境较嘈杂的情况中，可在训练样本集中包括较大比例的扰动样本，例如在训练样本集中包括全部扰动样本中的 80% 的扰动样本，全部初始样本中的 20% 的初始样本。相应地，也可以通过同样的比例配置测试样本集，例如包括全部扰动样本中剩余的（占全部扰动样本）20% 的扰动样本、剩余初始样本中的（占全部初始样本）5% 的初始样本。

[86] 在如图 5 和图 6 所示获取模型的训练样本集和测试样本集之后，可通过训练样本集训练模型，并可通过测试样本集评估对模型进行评估。下面描述基于测试样本集对模型的评估挑选模型的预定参数 λ ，从而进一步优化模型的方法。

[87] 图 7 示出根据本说明书实施例的一种模型训练方法流程图，包括：

[88] 在步骤 S702，获取初始样本集，其中，所述初始样本集中包括多个初始样本；

[89] 在步骤 S704, 通过图 6 所示方法基于所述初始样本集获取多个训练样本集、及与
所述多个训练样本集分别对应的多个测试样本集, 其中, 所述多个训练样本集与多个取
值不同的第一参数分别对应;

5 [90] 在步骤 S706, 使用所述多个训练样本集分别训练当前模型, 以分别获取多个更新
模型;

[91] 在步骤 S708, 使用所述多个测试样本集分别评估相应的更新模型, 其中, 所述测
试样本集与相应的更新模型对应于相同的训练样本集; 以及

[92] 在步骤 S710, 基于评估结果, 在所述多个更新模型中确定所述当前模型的更新模
型。

10 [93] 首先, 在步骤 S702, 获取初始样本集, 其中, 所述初始样本集中包括多个初始样
本。所述模型不限于具体类型, 如上文所述, 其可以为监督学习模型、无监督学习模型、
以及强化学习模型中的任一类型。例如, 所述模型为如上文所述的语音识别模型, 其例
如为监督学习模型, 在该情况中, 可通过人工输入语音, 从该语音提取相应的特征向量,
从而使用所述特征向量和标签值(语义)作为模型的初始样本。然而, 该模型在实际应
15 用中可能会遇到不同的环境, 如安静的环境、多种具有不同噪声的嘈杂的环境等等。而
在单一环境下人工获取的初始样本无法模拟这么多不同的环境, 而在不同环境下人工获
取样本的成本也比较高。从而, 可通过该方法基于初始样本集进行样本的扩充, 从而获
取训练样本集。

20 [94] 在步骤 S704, 通过图 6 所示方法基于所述初始样本集获取多个训练样本集、及与
所述多个训练样本集分别对应的多个测试样本集, 其中, 所述多个训练样本集与多个取
值不同的第一参数分别对应。

[95] 也就是说, 在对上述预定参数 λ 取不同值的情况中, 多次通过图 6 所示方法基于初
始样本集获取多个训练样本集和与其分别相应的多个测试样本集。例如, 可以将 λ 分别
取值为 0.0001、0.001、0.01、0.1。从而可确定 λ 的量级对模型训练的影响。可以理解, λ 的取
25 值不限于上述方式和上述个数, 而是可以根据具体模型具体限定。具体是, 对于上述 4
个 λ 值, 基于上述初始样本集 A, 可通过图 2 所示方法分别获取 4 个扰动样本集 B_1 、 B_2 、
 B_3 、 B_4 , 假设, 基于该 4 个扰动样本集分别获取 4 组样本集合组 (C_1, D_1) 、 (C_2, D_2) 、
 (C_3, D_3) 、 (C_4, D_4) , 其中 C_i 表示训练样本集, D_i 表示测试样本集。

[96] 在步骤 S706, 使用所述多个训练样本集分别训练当前模型, 以分别获取多个更新

模型。

[97] 在上述实例中，也就是说，使用各个训练样本集 C_1 、 C_2 、 C_3 、 C_4 分别训练当前模型，以分别获取 4 个更新模型 M_1 、 M_2 、 M_3 、 M_4 。

5 [98] 在步骤 S708，使用所述多个测试样本集分别评估相应的更新模型，其中，所述测试样本集与相应的更新模型对应于相同的训练样本集。

[99] 在上述实例中，也就是说，使用各个测试样本集 D_1 、 D_2 、 D_3 、 D_4 分别评估所述 4 个更新模型 M_1 、 M_2 、 M_3 、 M_4 ，其中，测试样本集 D_1 与更新模型 M_1 都对应于训练样本集 C_1 ，即，测试样本集 D_1 与更新模型 M_1 相对应，类似地可得出，测试样本集 D_2 与更新模型 M_2 相对应，测试样本集 D_3 与更新模型 M_3 相对应，测试样本集 D_4 与更新模型
10 M_4 相对应。可通过测试样本计算相应的更新模型的各种评估指标，如准确率、精确率、召回率等等，从而对该相应的更新模型进行评估，例如，可综合上述各种评估指标获取该模型的评估值。

[100] 在步骤 S710，基于评估结果，在所述多个更新模型中确定所述当前模型的更新模型。

15 [101] 在上述实例中，在获取各个 λ 对应的更新模型各自的评估值之后，例如可将评估值最高的更新模型确定为所述当前模型的更新模型，即训练后模型，并保留该确定的更新模型以进行后续的模式使用，如模型预测等。

[102] 图 8 示出根据本说明书实施例的一种基于初始样本集获取扰动样本集的装置 800，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述装置包
20 括：

[103] 计算单元 81，配置为，计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

[104] 生成单元 82，配置为，对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的
25 的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

[105] 图 9 示出根据本说明书实施例的一种基于初始样本集获取模型训练样本集的装置 900，其中，所述初始样本集中包括多个初始样本，所述装置包括：

[106] 获取单元 91, 配置为, 通过上述装置, 基于所述初始样本集获取扰动样本集, 所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本; 以及

[107] 合并单元 92, 配置为, 通过将所述多个初始样本与所述多个扰动样本合并, 获取训练样本集。

5 [108] 图 10 示出根据本说明书实施例的一种基于初始样本集获取模型训练样本集和测试样本集的装置 1000, 其中, 所述初始样本集中包括多个初始样本, 所述装置包括:

[109] 获取单元 101, 配置为, 通过上述获取扰动样本集的装置, 基于所述初始样本集获取扰动样本集, 所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本;

10 [110] 第一合并单元 102, 配置为, 通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并, 获取训练样本集; 以及

[111] 第二合并单元 103, 配置为, 通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并, 获取测试样本集。

[112] 在一个实施例中, 所述第二合并单元还配置为, 通过将所述多个初始样本中剩余的初始样本与所述多个扰动样本中剩余的扰动样本合并, 获取测试样本集。

15 [113] 图 11 示出根据本说明书实施例的一种模型训练装置 1100, 包括:

[114] 第一获取单元 111, 配置为, 获取初始样本集, 其中, 所述初始样本集中包括多个初始样本;

20 [115] 第二获取单元 112, 配置为, 通过上述获取训练样本集和测试样本集的装置基于所述初始样本集获取多个训练样本集、及与所述多个训练样本集分别对应的多个测试样本集, 其中, 所述多个训练样本集与多个取值不同的第一参数分别对应;

[116] 训练单元 113, 配置为, 使用所述多个训练样本集分别训练当前模型, 以分别获取多个更新模型;

[117] 评估单元 114, 配置为, 使用所述多个测试样本集分别评估相应的更新模型, 其中, 所述测试样本集与相应的更新模型对应于相同的训练样本集; 以及

25 [118] 确定单元 115, 配置为, 基于评估结果, 在所述多个更新模型中确定所述当前模型的更新模型。

[119] 本说明书另一方面提供一种计算机可读存储介质, 其上存储有计算机程序, 当所述计算机程序在计算机中执行时, 令计算机执行上述任一项方法。

[120] 本说明书另一方面提供一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现上述任一项方法。

[121] 本说明书实施例通过对模型的训练数据进行扰动，模拟真实环境中的数据噪声，从而增加模型的鲁棒性，并且通过使用扰动数据训练并评估模型，并基于评估结果确定模型预定参数，从而定量地提高了模型对异常数据的有效性，另外，本说明书实施例中未对机器学习模型的参数进行限制，从而不会限制模型的学习能力。

[122] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[123] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[124] 本领域普通技术人员应该还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执轨道，取决于技术方案的特定应用和设计约束条件。本领域普通技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

[125] 结合本文中所公开的实施例描述的方法或算法的步骤可以用硬件、处理器执轨道的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[126] 以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求书

1.一种基于初始样本集获取扰动样本集的方法，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述方法包括：

5 计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值均方差的乘积确定。

2.根据权利要求1所述的方法，其中，所述随机数为高斯分布随机数，所述高斯分布随机数的均方差为所述第一参数与该随机数对应的维度的特征值均方差的乘积。

3.根据权利要求1所述的方法，其中，所述随机数为平均随机数，其中，所述平均随机数的取值范围在正负第一数值之间，其中，所述第一数值为所述第一参数与该随机数对应的维度的特征值均方差的乘积。

4.一种基于初始样本集获取模型训练样本集的方法，其中，所述初始样本集中包括多个初始样本，所述方法包括：

通过权利要求1所述的方法，基于所述初始样本集获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；以及

20 通过将所述多个初始样本与所述多个扰动样本合并，获取训练样本集。

5.一种基于初始样本集获取模型训练样本集和测试样本集的方法，其中，所述初始样本集中包括多个初始样本，所述方法包括：

通过权利要求1所述的方法，基于所述初始样本集获取扰动样本集，所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本；

25 通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并，获取训练样本集；以及

通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集。

6.根据权利要求5所述的方法，其中，所述部分初始样本占所述多个初始样本的比例与所述部分扰动样本占所述多个扰动样本的比例相同。

7.根据权利要求6所述的方法，其中，通过将所述多个初始样本中剩余的初始样本

的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并，获取测试样本集包括，通过将所述多个初始样本中剩余的初始样本与所述多个扰动样本中剩余的扰动样本合并，获取测试样本集。

5 8.根据权利要求 6 所述的方法，其中，所述部分初始样本与所述部分扰动样本分别对应。

9.一种模型训练方法，包括：

获取初始样本集，其中，所述初始样本集中包括多个初始样本；

10 通过权利要求 5 所述的方法基于所述初始样本集获取多个训练样本集、及与所述多个训练样本集分别对应的多个测试样本集，其中，所述多个训练样本集与多个取值不同的第一参数分别对应；

使用所述多个训练样本集分别训练当前模型，以分别获取多个更新模型；

使用所述多个测试样本集分别评估相应的更新模型，其中，所述测试样本集与相应的更新模型对应于相同的训练样本集；以及

基于评估结果，在所述多个更新模型中确定所述当前模型的更新模型。

15 10.根据权利要求 9 所述的方法，其中，所述模型为以下任一类模型：监督学习模型、无监督学习模型、和强化学习模型。

11.一种基于初始样本集获取扰动样本集的装置，所述初始样本集中包括多个初始样本，每个初始样本包括对应的特征向量，所述装置包括：

20 计算单元，配置为，计算所述多个初始样本分别对应的多个特征向量中每个维度的特征值的特征值均方差；以及

生成单元，配置为，对于所述多个特征向量中每个特征向量中的每个维度，生成相应的随机数，并将该特征向量的该维度的当前特征值更新为所述当前特征值与相应的随机数之和，以生成与所述多个特征向量分别对应的多个扰动样本，从而获取扰动样本集，其中，各个所述随机数的取值范围基于预定的第一参数与该随机数对应的维度的特征值
25 均方差的乘积确定。

12.根据权利要求 11 所述的装置，其中，所述随机数为高斯分布随机数，所述高斯分布随机数的均方差为所述第一参数与该随机数对应的维度的特征值均方差的乘积。

13.根据权利要求 11 所述的装置，其中，所述随机数为平均随机数，其中，所述平均随机数的取值范围在正负第一数值之间，其中，所述第一数值为所述第一参数与该随机数对应的维度的特征值均方差的乘积。
30

14.一种基于初始样本集获取模型训练样本集的装置，其中，所述初始样本集中包括

多个初始样本, 所述装置包括:

获取单元, 配置为, 通过权利要求 11 所述的装置, 基于所述初始样本集获取扰动样本集, 所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本; 以及

5 合并单元, 配置为, 通过将所述多个初始样本与所述多个扰动样本合并, 获取训练样本集。

15. 一种基于初始样本集获取模型训练样本集和测试样本集的装置, 其中, 所述初始样本集中包括多个初始样本, 所述装置包括:

获取单元, 配置为, 通过权利要求 11 所述的装置, 基于所述初始样本集获取扰动样本集, 所述扰动样本集包括与所述多个初始样本分别对应的多个扰动样本;

10 第一合并单元, 配置为, 通过将所述多个初始样本中的部分初始样本与所述多个扰动样本中的部分扰动样本合并, 获取训练样本集; 以及

第二合并单元, 配置为, 通过将所述多个初始样本中剩余的初始样本的至少部分与所述多个扰动样本中剩余的扰动样本的至少部分合并, 获取测试样本集。

16. 根据权利要求 15 所述的装置, 其中, 所述部分初始样本占所述多个初始样本的比例与
15 所述部分扰动样本占所述多个扰动样本的比例相同。

17. 根据权利要求 16 所述的装置, 其中, 所述第二合并单元还配置为, 通过将所述多个初始样本中剩余的初始样本与所述多个扰动样本中剩余的扰动样本合并, 获取测试样本集。

18. 根据权利要求 16 所述的装置, 其中, 所述部分初始样本与所述部分扰动样本分
20 别对应。

19. 一种模型训练装置, 包括:

第一获取单元, 配置为, 获取初始样本集, 其中, 所述初始样本集中包括多个初始样本;

25 第二获取单元, 配置为, 通过权利要求 15 所述的装置基于所述初始样本集获取多个训练样本集、及与所述多个训练样本集分别对应的多个测试样本集, 其中, 所述多个训练样本集与多个取值不同的第一参数分别对应;

训练单元, 配置为, 使用所述多个训练样本集分别训练当前模型, 以分别获取多个更新模型;

30 评估单元, 配置为, 使用所述多个测试样本集分别评估相应的更新模型, 其中, 所述测试样本集与相应的更新模型对应于相同的训练样本集; 以及

确定单元, 配置为, 基于评估结果, 在所述多个更新模型中确定所述当前模型的更

新模型。

20.根据权利要求 19 所述的装置，其中，所述模型为以下任一类模型：监督学习模型、无监督学习模型、和强化学习模型。

21.一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机
5 中执行时，令计算机执行权利要求 1-10 中任一项的所述的方法。

22.一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-10 中任一项所述的方法。

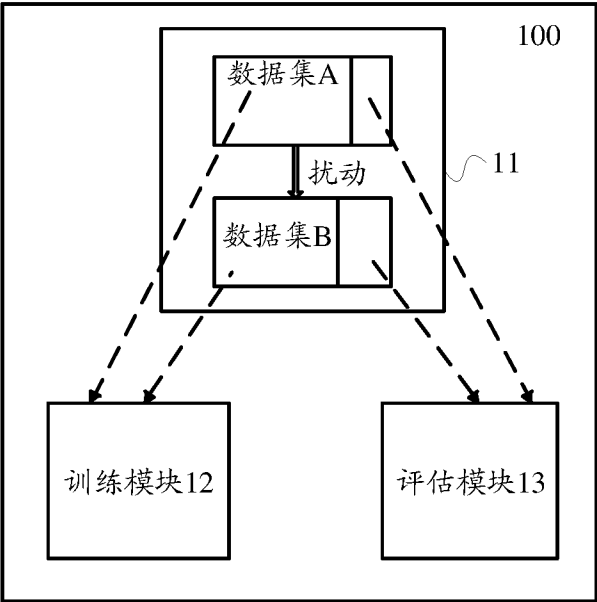


图 1

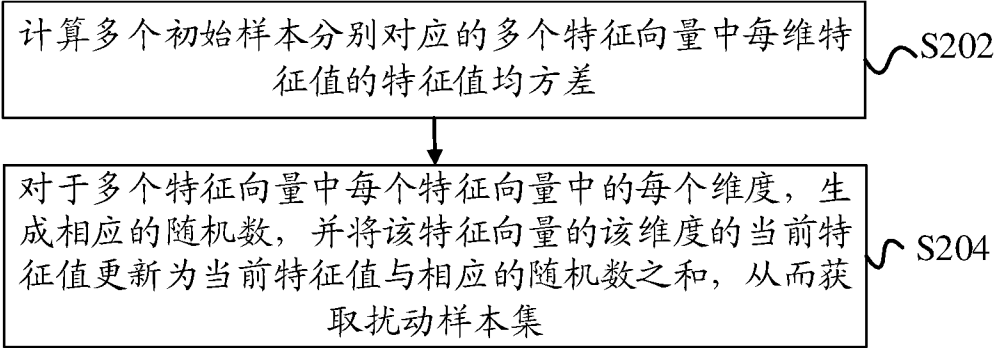


图 2

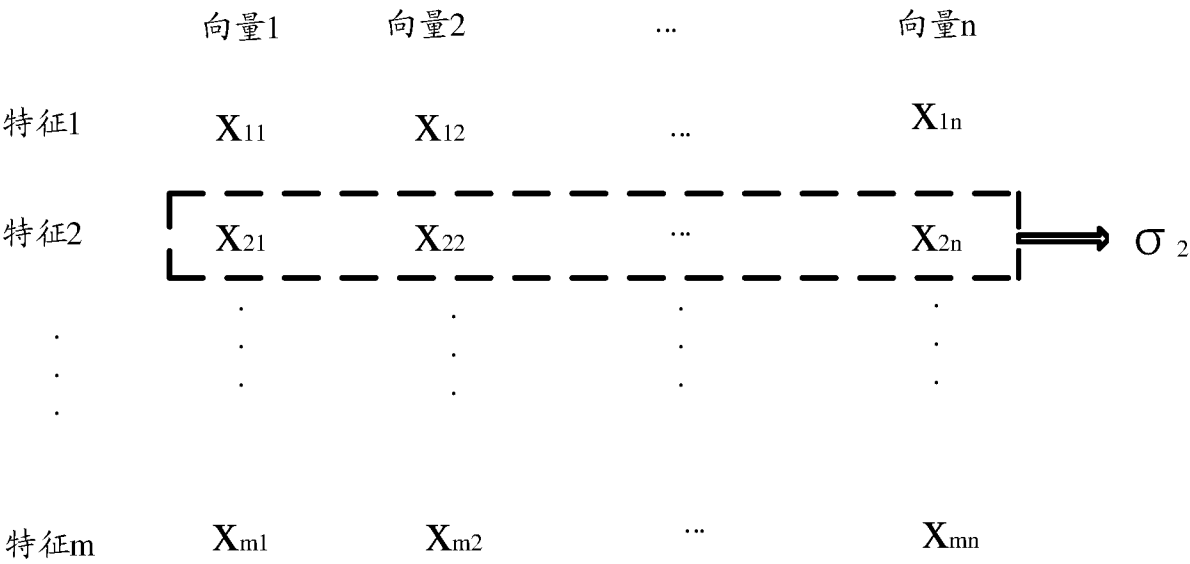


图 3

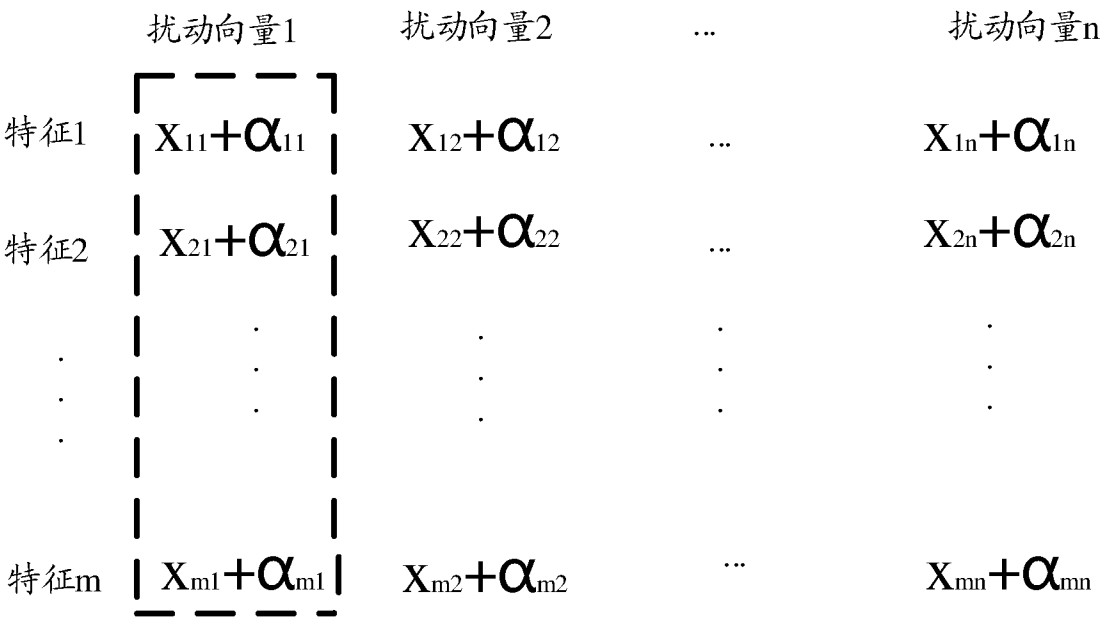


图 4

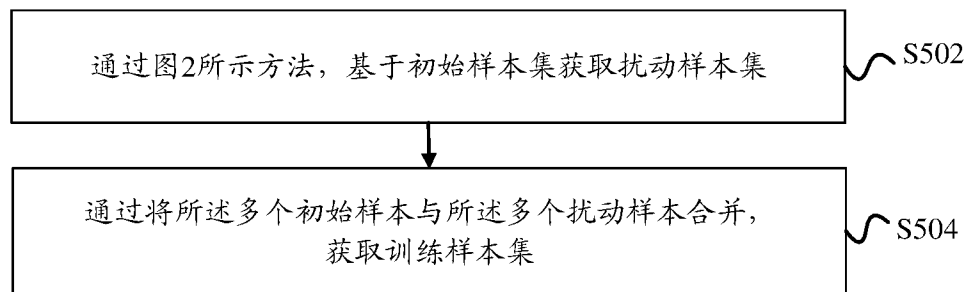


图 5

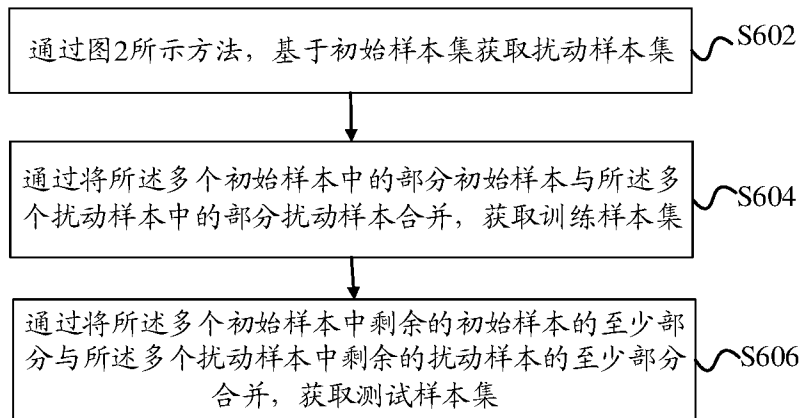


图 6

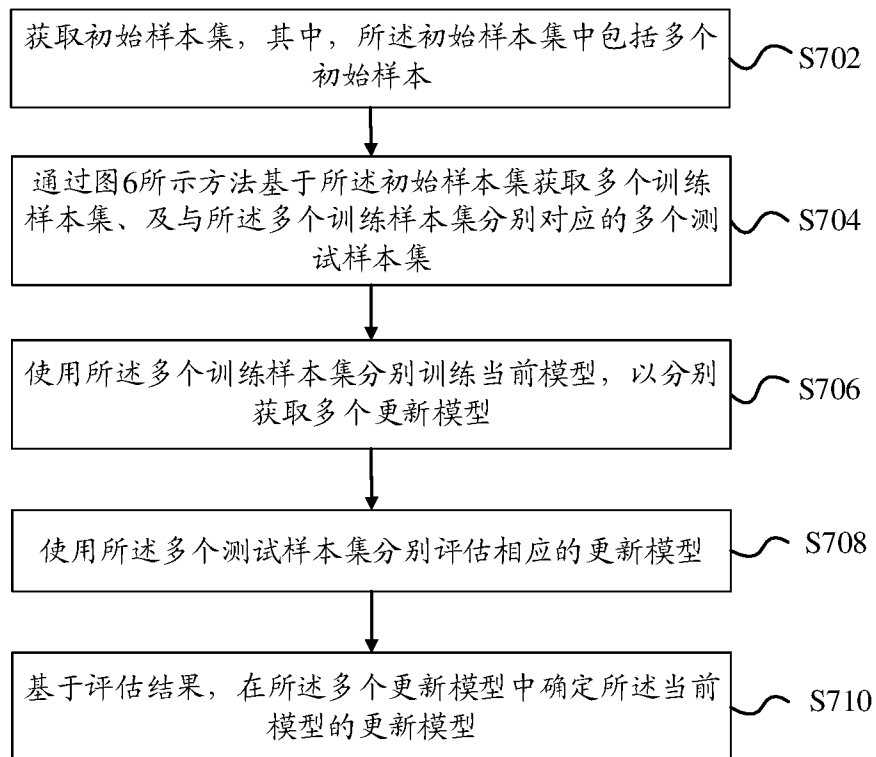


图 7

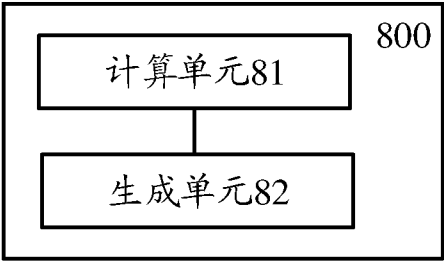


图 8

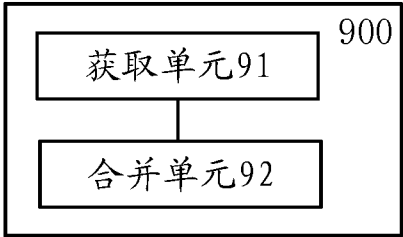


图 9

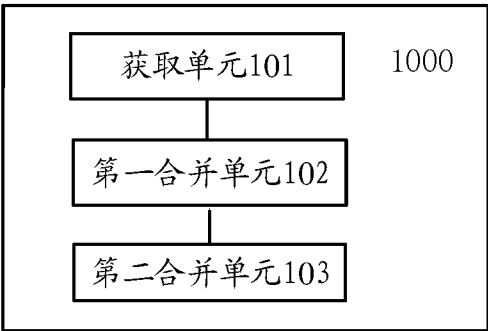


图 10

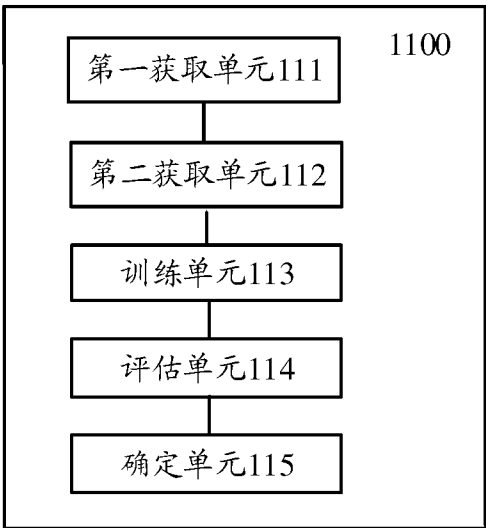


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070290

A. CLASSIFICATION OF SUBJECT MATTER G06N 20/00(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) G06N,G06F 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, WPI, EPODOC, CNKI: 扰, 噪声, 样本, 数据集, 加, 随机数, Disturb+, noise, sample, data, set, add, random number		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110033094 A (ALIBABA GROUP HOLDING LIMITED) 19 July 2019 (2019-07-19) claims 1-22	1-22
X	CN 107193863 A (GUANGDONG UNIVERSITY OF TECHNOLOGY) 22 September 2017 (2017-09-22) description, paragraphs [0007]-[0032] and [0048]-[0059]	1-8, 11-18, 21-22
Y	CN 107193863 A (GUANGDONG UNIVERSITY OF TECHNOLOGY) 22 September 2017 (2017-09-22) description, paragraphs [0007]-[0032] and [0048]-[0059]	9-10, 19-20
Y	CN 105808500 A (SHANXI PEONY ADVANCED INTELLIGENT TECHNOLOGY CO., LTD.) 27 July 2016 (2016-07-27) description, paragraphs 0050-0064	9-10, 19-20
A	CN 107315918 A (QINGDAO UNIVERSITY) 03 November 2017 (2017-11-03) entire document	1-22
A	US 2006048010 A1 (TAI, Hung'en et al.) 02 March 2006 (2006-03-02) entire document	1-22
☐ 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 21 March 2020		Date of mailing of the international search report 02 April 2020
Name and mailing address of the ISA/CN China National Intellectual Property Administration 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/CN2020/070290

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	110033094	A	19 July 2019	None	
CN	107193863	A	22 September 2017	None	
CN	105808500	A	27 July 2016	None	
CN	107315918	A	03 November 2017	None	
US	2006048010	A1	02 March 2006	None	

国际检索报告

国际申请号

PCT/CN2020/070290

A. 主题的分类 G06N 20/00 (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) G06N, G06F 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, WPI, EPDOC, CNKI: 扰, 噪声, 样本, 数据集, 加, 随机数, Disturb+, noise, sample, data, set, add, random number																							
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 110033094 A (阿里巴巴集团控股有限公司) 2019年 7月 19日 (2019 - 07 - 19) 权利要求1-22</td> <td>1-22</td> </tr> <tr> <td>X</td> <td>CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段</td> <td>1-8, 11-18, 21-22</td> </tr> <tr> <td>Y</td> <td>CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段</td> <td>9-10, 19-20</td> </tr> <tr> <td>Y</td> <td>CN 105808500 A (山西牡丹深度智能科技有限公司) 2016年 7月 27日 (2016 - 07 - 27) 说明书第0050-0064段</td> <td>9-10, 19-20</td> </tr> <tr> <td>A</td> <td>CN 107315918 A (青岛大学) 2017年 11月 3日 (2017 - 11 - 03) 全文</td> <td>1-22</td> </tr> <tr> <td>A</td> <td>US 2006048010 A1 (TAI, Hung-En 等) 2006年 3月 2日 (2006 - 03 - 02) 全文</td> <td>1-22</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 110033094 A (阿里巴巴集团控股有限公司) 2019年 7月 19日 (2019 - 07 - 19) 权利要求1-22	1-22	X	CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段	1-8, 11-18, 21-22	Y	CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段	9-10, 19-20	Y	CN 105808500 A (山西牡丹深度智能科技有限公司) 2016年 7月 27日 (2016 - 07 - 27) 说明书第0050-0064段	9-10, 19-20	A	CN 107315918 A (青岛大学) 2017年 11月 3日 (2017 - 11 - 03) 全文	1-22	A	US 2006048010 A1 (TAI, Hung-En 等) 2006年 3月 2日 (2006 - 03 - 02) 全文	1-22
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 110033094 A (阿里巴巴集团控股有限公司) 2019年 7月 19日 (2019 - 07 - 19) 权利要求1-22	1-22																					
X	CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段	1-8, 11-18, 21-22																					
Y	CN 107193863 A (广东工业大学) 2017年 9月 22日 (2017 - 09 - 22) 说明书第[0007]-[0032], [0048]-[0059]段	9-10, 19-20																					
Y	CN 105808500 A (山西牡丹深度智能科技有限公司) 2016年 7月 27日 (2016 - 07 - 27) 说明书第0050-0064段	9-10, 19-20																					
A	CN 107315918 A (青岛大学) 2017年 11月 3日 (2017 - 11 - 03) 全文	1-22																					
A	US 2006048010 A1 (TAI, Hung-En 等) 2006年 3月 2日 (2006 - 03 - 02) 全文	1-22																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 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 </td> <td> “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 </td> </tr> </table>			* 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																			
* 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 3月 21日		国际检索报告邮寄日期 2020年 4月 2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		授权官员 胡丽丽 电话号码 86-(10)-53961386																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070290

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	110033094	A	2019年 7月 19日	无	
CN	107193863	A	2017年 9月 22日	无	
CN	105808500	A	2016年 7月 27日	无	
CN	107315918	A	2017年 11月 3日	无	
US	2006048010	A1	2006年 3月 2日	无	