

(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



(10) 国际公布号

WO 2021/189980 A1

(43) 国际公布日

2021 年 9 月 30 日 (30.09.2021)

(51) 国际专利分类号:

G10L 15/16 (2006.01)

(21) 国际申请号:

PCT/CN2020/136366

(22) 国际申请日:

2020 年 12 月 15 日 (15.12.2020)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202011153538.8 2020 年 10 月 26 日 (26.10.2020) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];

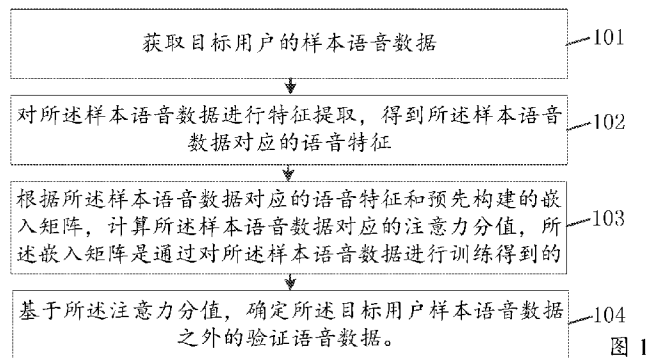
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 罗剑(LUO, Jian); 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)。王健宗(WANG, Jianzong); 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)。程宁(CHENG, Ning); 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: 深圳市明日今典知识产权代理有限公司(普通合伙)(SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM(GENERAL)); 中国广东省深圳市南山区南头街道安乐社区关口二路 15 号智恒产业园 30 栋 4 层 406 室, Guangdong 518000 (CN)。

(54) Title: VOICE DATA GENERATION METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 语音数据生成方法、装置、计算机设备及存储介质



- 101 Acquire sample voice data of a target user
- 102 Perform feature extraction on the sample voice data to obtain a voice feature corresponding to the sample voice data
- 103 According to the voice feature corresponding to the sample voice data and a pre-constructed embedding matrix, calculate an attention score corresponding to the sample voice data, wherein the embedding matrix is obtained by training the sample voice data
- 104 On the basis of the attention score, determine verification voice data other than the sample voice data of the target user

(57) Abstract: A voice data generation method and apparatus, and a computer device and a storage medium, which relate to the technical field of artificial intelligence and are mainly characterized in that more voice data of a few sample users can be generated on the basis of sample voice data of these sample users, thereby balancing voice data of different users in a sample library. The method comprises: acquiring sample voice data of a target user (101); performing feature extraction on the sample voice data to obtain a voice feature corresponding to the sample voice data (102); according to the voice feature corresponding to the sample voice data and a

(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, IT, 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, 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))。
- 在修改权利要求的期限届满之前进行, 在收到该修改后将重新公布(细则48.2(h))。
- 根据申请人的请求, 在条约第21条(2)(a)所规定的期限届满之前进行。

pre-constructed embedding matrix, calculating an attention score corresponding to the sample voice data, wherein the embedding matrix is obtained by training the sample voice data (103); and on the basis of the attention score, determining verification voice data other than the sample voice data of the target user (104). In the method, machine learning technology is used, and the method is mainly applicable to the generation of voice data.

(57) 摘要: 一种语音数据生成方法、装置、计算机设备及存储介质, 涉及人工智能技术领域, 主要在于能够基于少数样本用户的样本语音数据, 生成少数样本用户更多的语音数据, 从而使得样本库中的不同用户的语音数据达到平衡。其中方法包括: 获取目标用户的样本语音数据(101); 对所述样本语音数据进行特征提取, 得到所述样本语音数据对应的语音特征(102); 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵, 计算所述样本语音数据对应的注意力分值, 所述嵌入矩阵是通过所述样本语音数据进行训练得到的(103); 基于所述注意力分值, 确定所述目标用户样本语音数据之外的验证语音数据(104)。本方法采用了机器学习技术, 主要适用于语音数据的生成。

语音数据生成方法、装置、计算机设备及存储介质

本申请要求于2020年10月26日提交中国专利局、申请号为202011153538.8,发明名称为“语音数据生成方法、装置、计算机设备及存储介质”的中国专利申请的优先权,其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及人工智能技术领域,尤其是涉及一种语音数据生成方法、装置、计算机设备及存储介质。

10 背景技术

在语音识别技术中,对于分类器来说,如果样本库中不同用户的语音数据量差距过大,分类器很难有良好的表现,会影响分类器识别的准确率,因此,为了确保语音识别的准确率,需要保证样本语音数据的平衡。

15 目前,对于不平衡的样本语音数据,通常采用欠采样的方式对多数样本语音数据进行数据消除,或者人为采集少数样本用户的语音数据进行数据补充。然而,发明人意识到如果消除多数样本语音数据中的部分数据,很可能会丢失有价值的用户信息,进而影响用户的语音识别精度,此外,对于补充语音数据的方式,由于受到用户隐私和安全因素的限制,很难获得少数样本用户的大量语音数据,且这种人为采集语音数据的方式,操作较为不便。

20 技术问题

本申请提供了一种语音数据生成方法、装置、计算机设备及存储介质,主要在于解决如何使得样本库中的不同用户的语音数据达到平衡,同时能够避免丢失有价值的用户信息的问题。

25 技术解决方案

根据本申请的第一个方面,提供一种语音数据生成方法,包括:

获取目标用户的样本语音数据;

对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征;

30 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过所述样本语音数据进行训练得到的;

基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据。

根据本申请的第二个方面,提供一种语音数据生成装置,包括:

获取单元,用于获取目标用户的样本语音数据;

35 提取单元,用于对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征;

第一确定单元,用于根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过所述样本语音数据进行训练得到的,

40 第二确定单元,用于基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据。

根据本申请的第三个方面,提供一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现语音数据生成方法;

其中,所述语音数据生成方法的步骤包括:

获取目标用户的样本语音数据；

对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；

根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过所述样本语音数据进行训练得到的；

5 基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。

根据本申请的第四个方面，提供一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现语音数据生成方法；

其中，所述语音数据生成方法的步骤包括：

10 获取目标用户的样本语音数据；

对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；

根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过所述样本语音数据进行训练得到的；

基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。

15

有益效果

本申请提供了一种语音数据生成方法、装置、计算机设备及存储介质，通过提取样本数据量匮乏的目标用户的语音特征，能够计算样本语音数据对应的注意力分值，并依据该注意力分值生成目标用户的验证语音数据，从而能够根据目标用户的少数样本语音数据，生成更多的语音数据，使不同用户的样本语音数据达到平衡，避免采用欠采用的方式丢失掉有价值的用户信息，同时依据该样本语音数据训练的预设语音识别模型的语音识别精度也得到了提高。

20

附图说明

25 此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

图 1 示出了本申请实施例提供的一种语音数据生成方法流程图；

图 2 示出了本申请实施例提供的另一种语音数据生成方法流程图；

图 3 示出了本申请实施例提供的一种语音数据生成装置的结构示意图；

30 图 4 示出了本申请实施例提供的另一种语音数据生成装置的结构示意图；

图 5 示出了本申请实施例提供的一种计算机设备的实体结构示意图。

本发明的最佳实施方式

35 下文中将参考附图并结合实施例来详细说明本申请。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

目前，对于不平衡的样本语音数据，通常采用欠采样的方式对多数样本语音数据进行数据消除，或者人为采集少数样本用户的语音数据进行数据补充。然而，如果消除多数样本语音数据中的部分数据，很可能会丢失有价值的用户信息，进而影响用户的语音识别精度，此外，对于补充语音数据的方式，由于受到用户隐私和安全因素的限制，很难获得少数样本用户的大量语音数据，且这种人为采集语音数据的方式，操作较为不便。

40

为了解决上述问题，本申请实施例提供了一种语音数据生成方法，如图 1 所示，所述方法包括：

101、获取目标用户的样本语音数据。

其中，目标用户为样本语音数据匮乏的用户，该目标用户的样本语音数据的数据量小

于预设数据量，目标用户的样本语音数据为预设样本库中已经存在的语音数据，具体可以通过搜集不同用户的语音数据，构建预设样本库，在语音搜集的过程中，可能由于隐私和安全因素的限制，有些用户的样本语音数据量相比其他用户较少，为了确保训练的语音识别模型的语音识别精度，需要保证预设样本库不同用户的语音数据达到平衡，因此需要利用目标用户已有的样本语音数据，生成更多的样本语音数据，以达到预设样本库中不同用户的样本语音数据的平衡，确保后续的预设语音识别模型的语音识别精度，本申请实施例主要适用于语音数据的生成，本申请实施例的执行主体为能够生成目标语音数据的装置或设备，具体可以设置在客户端或者服务器一侧。

对于本申请实施例，为了筛选预设样本库中的目标用户，首先确定预设样本库各个用户对应的语音数据，基于各个用户对应的语音数据，统计各个用户对应的语音数据量，之后根据各个用户对应的语音数据量，统计预设样本库的语音数据量均值，将该语音数据量均值确定为预设数据量，接着分别将各个用户对应的语音数据量与预设数据量进行比对，并将语音数据量小于预设数据量的用户确定为目标用户，若用户的语音数据量小于预设数据量说明，该用户的语音数据量与预设样本库中其他用户的语音数据量相比较少，因此将语音数据量小于预设数据量的用户确定为目标用户，同时将预设样本库中目标用户的语音数据确定为目标用户的样本语音数据，以便根据该样本语音数据，生成目标用户除样本语音数据之后的新增样本数据，以到达样本语音数据的平衡。

102、对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征。

其中，可以将样本语音数据对应的梅尔倒谱系数作为样本语音数据对应的语音特征，具体地，在对样本语音数据进行特征提取之前需要对样本语音数据进行预处理，该预处理过程具体包括预加重、分帧和加窗函数处理，从而使得目标用户的样本语音数据变得平坦，即将样本语音数据的每N个采用点合成一个观测单位（帧），帧的左右端具有连续性，在对目标用户的样本语音数据进行预处理之后，需要对预处理后的样本语音数据进行快速傅里叶转化，得到转换后的语音数据，之后将转换后的语音数据输入Mel滤波器，计算转换后的语音数据通过Mel滤波器后的语音能量，接着根据样本语音数据对应的语音能量，计算样本语音数据对应的梅尔倒谱系数，并将该梅尔倒谱系数确定为目标用户的样本语音数据对应的语音特征，以便依据样本语音数据对应的语音特征，生成目标用户更多的语音数据，以便到预设样本库中样本语音数据的平衡。

103、根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值。

其中，所述嵌入矩阵是通过对所述样本语音数据进行训练得到的，对于本申请实施例，为了基于目标用户的样本语音数据，生成目标用户样本语音数据之外的验证语音数据，可以利用预设语音数据生成模型来生成更多目标用户的语音数据，由于目标用户的语音数据为时序数据，而GPT-2模型能够很好的处理时序数据，因此预设语音数据生成模型具体可以为预设GPT-2模型，该预设GPT-2模型中包括注意力层和神经网络层，具体地，将提取的样本语音数据对应的语音特征输入至预设GPT-2模型中进行语音数据生成，在预设GPT-2模型进行语音数据生成的过程中，首先将样本语音数据对应的语音特征输入至注意力层，利用该注意力层计算已有语音特征对应的注意力分值，具注意力层具体计算语音特征对应的注意力分值时，可以获取训练好的GPT-2模型中的嵌入矩阵，之后根据该嵌入矩阵计算该语音特征对应的查询向量、键向量和值向量，接着根据计算的查询向量、键向量和值向量，计算语音特征对应的注意力分值。

104、基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。

对于本申请实施例，将计算出的语音特征对应的注意力分值输入至神经网络层进行语音数据的生成，具体地，目标用户通常具有多个语音特征，在确定多个语音特征对应的注意力分值之后，将注意力分值输入至神经网络层，该神经网络层会筛选注意力分值较高的

语音特征, 语音特征的注意力分值越高, 说明该语音特征与待生成语音数据的关联性越高, 进而利用注意力分值较高的语音特征来生成目标用户的语音数据。

本申请实施例提供的一种语音数据生成方法, 与目前采用欠采样的方式对多数样本语音数据进行数据消除的方式相比, 本申请能够获取目标用户的样本语音数据; 并对所述样本语音数据进行特征提取, 得到所述样本语音数据对应的语音特征; 与此同时, 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵, 计算所述样本语音数据对应的注意力分值, 所述嵌入矩阵是通过对所述样本语音数据进行训练得到的; 并基于所述注意力分值, 确定所述目标用户样本语音数据之外的验证语音数据。由此通过提取样本数据量匮乏的目标用户的语音特征, 能够计算样本语音数据对应的注意力分值, 并依据该注意力分值生成目标用户的验证语音数据, 从而能够根据目标用户的少数样本语音数据, 生成更多的语音数据, 使不同用户的样本语音数据达到平衡, 避免采用欠采用的方式丢失掉有价值的用户信息, 同时依据该样本语音数据训练的预设语音识别模型的语音识别精度也得到了提高。

进一步的, 为了更好的说明上述语音数据的生成过程, 作为对上述实施例的细化和扩展, 本申请实施例提供了另一种语音数据生成方法, 如图 2 所示, 所述方法包括:

201、获取目标用户的样本语音数据。

其中, 样本语音数据为预设样本库中已经存在的语音数据, 目标用户为预设样本库中样本语音数据匮乏的用户, 对于本申请实施例, 为了确定目标用户, 可以预先设定预设语音数据量, 该预设语音数据量具体可以根据构建预设语音数据生成模型所需的训练样本量进行确定, 之后确定预设样本库中各个用户对应的语音数据量, 并将各个用户对应的语音数据量分别与预设语音数据量进行对比, 根据对比结果筛选目标用户, 具体可以将语音数据量小于预设语音数据量的用户确定为目标用户, 此外, 还可以根据各个用户对应的语音数据量, 计算预设样本库的语音数据量均值, 并将各个用户对应的语音数据量分别与语音数据量均值进行对比, 根据对比结果筛选目标用户, 具体可以将语音数据量小于语音数据量均值的用户确定为目标用户, 由此能够确定预设样本库中数据量匮乏的目标用户, 以便根据目标用户的样本语音数据, 生成目标用户更多的语音数据, 以达到预设样本库中语音数据的平衡。

202、对所述样本语音数据进行特征提取, 得到所述样本语音数据对应的语音特征。

其中, 样本语音数据对应的语音特征具体可以为样本语音数据对应的梅尔倒谱系数, 对于本申请实施例, 为了提取样本语音数据对应的语音特征, 步骤 202 具体包括: 对所述样本语音数据进行滤波处理, 得到所述样本语音数据对应的语音能量; 对所述语音能量进行离散余弦化处理, 得到所述样本语音数据对应的语音特征。

具体地, 在对样本语音数据进行特征提取之前需要对样本语音数据进行预处理, 该预处理过程具体包括预加重、分帧和加窗函数处理, 从而使得目标用户的样本语音数据变得平坦, 即将样本语音数据的每 N 个采用点合成一个观测单位 (帧), 帧的左右端具有连续性, 在对目标用户的样本语音数据进行预处理之后, 需要对预处理后的样本语音数据进行快速傅里叶转化, 得到转换后的语音数据, 之后将转换后的语音数据输入 Mel 滤波器, 计算转换后的语音数据通过 Mel 滤波器后的语音能量, 接着根据样本语音数据对应的语音能量, 计算样本语音数据对应的梅尔倒谱系数, 并将该梅尔倒谱系数确定为目标用户的样本语音数据对应的语音特征, 梅尔倒谱系数的具体计算公式如下:

$$C(n) = \sum_{m=1}^{N-1} s(m) * \cos\left(\frac{\pi n (m-5)}{M}\right), n = 1, 2, \dots, L$$

其中, $s(m)$ 代表语音数据经过第 m 个滤波器后输出的语音能量, M 为滤波器的总个数, $C(n)$ 为梅尔倒谱系数, n 代表梅尔倒谱系数的阶数, L 通常可取 12-16, $s(m)$ 语音能量的具

体计算公式如下:

$$s(m) = \ln\left(\sum_{K=0}^{N-1} |X_{\phi}(k)|^2 H_m(k)\right), 0 \leq m \leq M$$

其中, $X_{\phi}(k)$ 为对语音数据的频谱取模平方得到语音数据的功率谱, $H_m(k)$ 为滤波器的频率, K 为傅里叶变换的点数。由此按照上述公式, 能够计算出目标用户样本语音数据对应的梅尔倒谱系数, 并将其确定为样本语音数据对应的语音特征, 以便根据该样本语音数据, 生成目标用户除样本语音数据之外的新增样本数据, 以到达样本语音数据的平衡。

203、根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵, 计算所述样本语音数据对应的注意力分值。

其中, 所述嵌入矩阵是通过对所述样本语音数据进行训练得到的, 对于本申请实施例, 为了获取目标用户更多的语音数据, 以达到样本语音数据的平衡, 将样本语音数据对应的语音特征输入至预设语音数据生成模型进行数据生成, 得到目标用户样本语音数据之外的验证语音数据, 其中, 预设语音数据生成模型具体可以为已经训练好的 GPT-2 模型, 具体利用 GPT-2 模型生成目标用户更多的语音数据时, 步骤 203 具体包括: 根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量; 将所述语音特征对应的查询向量和与其对应的键向量相乘, 得到所述语音特征对应的权重值; 根据所述语音特征对应的权重值和值向量, 计算所述语音特征对应的注意力分值。

具体地, 预设嵌入矩阵是由训练好的 GPT-2 模型确定的, 即通过训练 GPT-2 模型能够得到预设嵌入矩阵, 之后根据该预设嵌入矩阵能够确定语音特征对应的查询向量、键向量和值向量, 之后在 GPT-2 模型中的注意力层根据语音特征对应的查询向量和键向量, 计算语音特征对应的权重值, 接着根据语音特征对应的权重值和值向量, 计算语音特征对应的注意力分值并输出, 该注意力分值的具体计算公式如下:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V$$

其中, $\text{Attention}(Q, K, V)$ 为已有特征对应的注意力分值, Q 为查询向量, K 为键向量, V 为值向量, d_K 为键向量的维数, 通常取 64。由此能够得到语音特征对应的注意力得分,

204、基于所述注意力分值, 确定所述目标用户样本语音数据之外的验证语音数据。

对于本申请实施例, 将该注意力得分输入至 GPT-2 模型中的神经网络层, 生成目标用户除样本语音数据之外的验证语音数据, 以确保样本库中不同用户的语音数据量达到平衡。

205、将所述目标用户的样本语音数据和验证语音数据, 以及预设样本库中其他用户的样本语音数据确定为第一训练样本。

其中, 其他用户为语音数据量不匮乏的用户, 即其他用户对应的语音数据量大于预设数据量, 对于本方实施例, 生成目标用户更多的语音数据后, 样本库中不同用户的语音数据量达到平衡, 可以将样本库中的语音数据作为训练样本, 构建预设语音识别模型, 具体地, 将目标用户的样本语音数据和验证语音数据, 以及预设样本库中其他用户的样本语音数据共同作为第一训练样本, 以便根据该第一训练样本构建预设语音识别模型。

206、利用预设神经网络算法对所述第一训练样本进行训练, 构建预设语音识别模型。

对于本申请实施例, 预设语音识别模型具体可以为预设神经网络模型, 该预设神经网络模型包括多个隐藏层, 给定预设神经网络模型的初始参数, 之后将第一训练样本输入至预设神经网络模型中进行训练, 即对预设神经网络模型中的初始参数进行调整, 构建预设语音识别模型。

进一步地, 为了确保预设语音识别模型能够对真实的语音数据有更好的识别效果, 还

可以利用目标用户的样本语音数据以及预设样本库中其他用户的语音数据对构建的预设语音识别模型进行调整,基于此,所述方法还包括:

将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本;利用所述第二训练样本对所述预设语音识别模型进行调整,得到调整后的预设语音识别模型。由此调整后的预设语音识别模型能够对真实的语音数据有更好的识别效果。

进一步地,为了保证调整的预设语音识别模型的识别精度,所述方法还包括:利用测试样本对所述调整后的预设语音识别模型进行测试,得到所述调整后的预设语音识别模型对应的测试结果;根据所述测试结果,确定所述调整后的预设语音识别模型对应的语音识别准确率;若所述语音识别准确率小于预设语音识别准确率,对所述调整后的预设语音识别模型中的参数进行调整,直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。具体地,获取多个用户的测试样本,将测试样本输入至调整后的预设语音识别模型进行测试,能够得到调整后的预设语音识别模型的测试结果,根据该测试结果,统计测试样本中识别结果正确的样本数量和样本总数,并根据识别结果正确的样本数量和样本总数,计算调整后的预设语音识别模型对应的语音识别准确率,如果计算的语音识别准确率未达到预设语音识别准确率,则确定调整后的预设语音识别模型的识别精度未达到要求,不可以进行语音识别,需要进行继续训练;如果计算的语音识别准确率达到预设语音识别准确率,则确定调整后的预设语音识别模型的识别精度得到要求,可以用来进行语音识别,基于此,所述方法还包括:获取待识别用户的语音数据;将所述待识别用户的语音数据输入至调整后的预设语音识别模型进行语音识别,确定所述待识别用户对应的语音识别结果。

具体地,将待识别用户的语音数据输入至调整后的预设语音识别模型进行语音识别,该调整后的预设语音识别模型中的隐藏层会提取待识别用户的语音数据对应的语音特征,并将待识别用户对应的语音特征与预设特征库中其他用户对应的语音特征进行比对,根据比对结果输出待识别用户对应的语音识别结果,即利用调整后的预设语音识别模型能够对待识别用户的身份进行识别。

本申请实施例提供的另一种语音数据生成方法,与目前采用欠采样的方式对多数样本语音数据进行数据消除的方式相比,本申请能够获取目标用户的样本语音数据;并对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征;与此同时,根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过所述样本语音数据进行训练得到的,并基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据。由此通过提取样本数据量匮乏的目标用户的语音特征,能够计算样本语音数据对应的注意力分值,并依据该注意力分值生成目标用户的验证语音数据,从而能够根据目标用户的少数样本语音数据,生成更多的语音数据,使不同用户的样本语音数据达到平衡,避免采用欠采用的方式丢失掉有价值的用户信息,同时依据该样本语音数据训练的预设语音识别模型的语音识别精度也得到了提高。

进一步地,作为图1的具体实现,本申请实施例提供了一种语音数据生成装置,如图3所示,所述装置包括:获取单元31、提取单元32、第一确定单元33和第二确定单元34。

所述获取单元31,可以用于获取目标用户的样本语音数据。所述获取单元31是本装置中获取目标用户的样本语音数据的主要功能模块。

所述提取单元32,可以用于对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征。所述提取单元32是本装置中对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征的主要功能模块,也是核心模块。

所述第一确定单元33,可以用于根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过所述样本语

音数据进行训练得到的。所述确定单元 33 是本装置中根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值的主要功能模块，也是核心模块。

所述第二确定单元 34，可以用于基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。所述第二确定单元是本装置中基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据的主要功能模块，也是核心模块。

进一步地，为了计算所述样本语音数据对应的注意力分值，如图 4 所示，所述第一确定单元 33，包括：确定模块 331、相乘模块 332 和计算模块 333。

所述确定模块 331，可以用于确定所述语音特征对应的嵌入矩阵，并根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量。

所述相乘模块 332，可以用于分别将所述语音特征对应的查询向量和与其对应的键向量相乘，得到所述语音特征对应的权重值。

所述计算模块 333，可以用于根据所述语音特征对应的权重值和值向量，计算所述语音特征对应的注意力分值。

进一步地，为了提取样本语音数据对应的语音特征，所述提取单元 32，包括：滤波模块 321 和离散模块 322。

所述滤波模块 321，可以用于对所述样本语音数据进行滤波处理，得到所述样本语音数据对应的语音能量。

所述离散模块 322，可以用于对所述语音能量进行离散余弦化处理，得到所述样本语音数据对应的语音特征。

进一步地，为了构建预设语音识别模型，所述装置还包括构建单元 35。

所述第一确定单元 33，还可以用于将所述目标用户的样本语音数据和验证语音数据，以及预设样本库中其他用户的样本语音数据确定为第一训练样本。

所述构建单元 35，可以用于利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型。

进一步地，为了提高预设语音识别模型的识别精度，所述装置还包括：调整单元 36。

所述第一确定单元 33，还可以用于将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本。

所述调整单元 36，可以用于利用所述第二训练样本对所述预设语音识别模型进行调整，得到调整后的预设语音识别模型。

进一步地，为了对调整后的预设语音识别模型进行测试，所述装置还包括测试单元 37。

所述测试单元 37，可以用于利用测试样本对所述调整后的预设语音识别模型进行测试，得到所述调整后的预设语音识别模型对应的测试结果。

所述第一确定单元 33，还可以用于根据所述测试结果，确定所述调整后的预设语音识别模型对应的语音识别准确率。

所述调整单元 36，还可以用于若所述语音识别准确率小于预设语音识别准确率，对所述调整后的预设语音识别模型中的参数进行调整，直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。

进一步地，为了对待识别用户进行语音识别，所述装置还包括：识别单元 38。

所述获取单元 31，还可以用于获取待识别用户的语音数据。

所述识别单元 38，可以用于将所述待识别用户的语音数据输入至调整后的预设语音识别模型进行语音识别，确定所述待识别用户对应的语音识别结果。

需要说明的是，本申请实施例提供的一种语音数据生成装置所涉及各功能模块的其他相应描述，可以参考图 1 所示方法的对应描述，在此不再赘述。

基于上述如图 1 所示方法，相应的，本申请实施例还提供了一种计算机可读存储介质，

所述计算机可读存储介质可以是非易失性，也可以是易失性，其上存储有计算机程序，该程序被处理器执行时实现语音数据生成方法；其中，所述语音数据生成方法的步骤包括：：获取目标用户的样本语音数据；对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算
5 所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过对所述样本语音数据进行训练得到的；基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。

进一步地，所述根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，包括：

根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量；

10 将所述语音特征对应的查询向量和键向量相乘，得到所述语音特征对应的权重值；

根据所述语音特征对应的权重值和值向量，计算所述语音特征对应的注意力分值。

进一步地，所述对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征，包括：

对所述样本语音数据进行滤波处理，得到所述样本语音数据对应的语音能量；

15 对所述语音能量进行离散余弦化处理，得到所述样本语音数据对应的语音特征。

进一步地，在所述基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据之后，所述方法还包括：

将所述目标用户的样本语音数据和验证语音数据，以及预设样本库中其他用户的样本语音数据确定为第一训练样本；

20 利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型。

进一步地，在所述利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型之后，所述方法还包括：

将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本；

25 利用所述第二训练样本对所述预设语音识别模型进行调整，得到调整后的预设语音识别模型。

进一步地，所述方法还包括：

利用测试样本对所述调整后的预设语音识别模型进行测试，得到所述调整后的预设语音识别模型对应的测试结果；

30 根据所述测试结果，确定所述调整后的预设语音识别模型对应的语音识别准确率；

若所述语音识别准确率小于预设语音识别准确率，对所述调整后的预设语音识别模型中的参数进行调整，直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。

基于上述如图 1 所示方法和如图 3 所示装置的实施例，本申请实施例还提供了一种计算机设备的实体结构图，如图 5 所示，该计算机设备包括：处理器 41、存储器 42、及存储在存储器 42 上并可在处理器上运行的计算机程序，其中存储器 42 和处理器 41 均设置在总线 43 上所述处理器 41 执行所述程序时实现语音数据生成方法；其中，所述语音数据生成方法的步骤包括：获取目标用户的样本语音数据；对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过对所述样本语音数据进行训练得到的；基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。
35

进一步地，所述根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，包括：

45 根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量；

将所述语音特征对应的查询向量和键向量相乘，得到所述语音特征对应的权重值；
根据所述语音特征对应的权重值和值向量，计算所述语音特征对应的注意力分值。

进一步地，所述对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征，包括：

5 对所述样本语音数据进行滤波处理，得到所述样本语音数据对应的语音能量；

对所述语音能量进行离散余弦化处理，得到所述样本语音数据对应的语音特征。

进一步地，在所述基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据之后，所述方法还包括：

10 将所述目标用户的样本语音数据和验证语音数据，以及预设样本库中其他用户的样本语音数据确定为第一训练样本；

利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型。

进一步地，在所述利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型之后，所述方法还包括：

15 将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本；

利用所述第二训练样本对所述预设语音识别模型进行调整，得到调整后的预设语音识别模型。

进一步地，所述方法还包括：

20 利用测试样本对所述调整后的预设语音识别模型进行测试，得到所述调整后的预设语音识别模型对应的测试结果；

根据所述测试结果，确定所述调整后的预设语音识别模型对应的语音识别准确率；

若所述语音识别准确率小于预设语音识别准确率，对所述调整后的预设语音识别模型中的参数进行调整，直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。

25 通过本申请的技术方案，本申请能够获取目标用户的样本语音数据；并对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；与此同时，根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过所述样本语音数据进行训练得到的；并基于所述注意力分值，
30 确定所述目标用户样本语音数据之外的验证语音数据，由此通过提取样本数据量匮乏的目标用户的语音特征，能够计算样本语音数据对应的注意力分值，并依据该注意力分值生成目标用户的验证语音数据，从而能够根据目标用户的少数样本语音数据，生成更多的语音数据，使不同用户的样本语音数据达到平衡，避免采用欠采用的方式丢失掉有价值的用户信息，同时依据该样本语音数据训练的预设语音识别模型的语音识别精度也得到了提高。

35 显然，本领域的技术人员应该明白，上述的本申请的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步
40 骤制作成单个集成电路模块来实现。这样，本申请不限制于任何特定的硬件和软件结合。

以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包括在本申请的保护范围之内。

权利要求书

- 1、一种语音数据生成方法，其中，包括：
获取目标用户的样本语音数据；
- 5 对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征；
 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，所述嵌入矩阵是通过对所述样本语音数据进行训练得到的；
 基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。
- 2、根据权利要求1所述的方法，其中，所述根据所述样本语音数据对应的语音特征
10 和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，包括：
 根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量；
 将所述语音特征对应的查询向量和键向量相乘，得到所述语音特征对应的权重值；
 根据所述语音特征对应的权重值和值向量，计算所述语音特征对应的注意力分值。
- 3、根据权利要求1所述的方法，其中，所述对所述样本语音数据进行特征提取，得
15 到所述样本语音数据对应的语音特征，包括：
 对所述样本语音数据进行滤波处理，得到所述样本语音数据对应的语音能量；
 对所述语音能量进行离散余弦化处理，得到所述样本语音数据对应的语音特征。
- 4、根据权利要求1所述的方法，其中，在所述基于所述注意力分值，确定所述目标
20 用户样本语音数据之外的验证语音数据之后，所述方法还包括：
 将所述目标用户的样本语音数据和验证语音数据，以及预设样本库中其他用户的样本
 语音数据确定为第一训练样本；
 利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型。
- 5、根据权利要求4所述的方法，其中，在所述利用预设神经网络算法对所述第一训
25 练样本进行训练，构建预设语音识别模型之后，所述方法还包括：
 将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样
 本；
 利用所述第二训练样本对所述预设语音识别模型进行调整，得到调整后的预设语音识
 别模型。
- 6、根据权利要求5所述的方法，其中，所述方法还包括：
30 利用测试样本对所述调整后的预设语音识别模型进行测试，得到所述调整后的预设语
 音识别模型对应的测试结果；
 根据所述测试结果，确定所述调整后的预设语音识别模型对应的语音识别准确率；
 若所述语音识别准确率小于预设语音识别准确率，对所述调整后的预设语音识别模型中
 的参数进行调整，直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语
35 音识别准确率。
- 7、根据权利要求5所述的方法，其中，所述方法还包括：
 获取待识别用户的语音数据；
 将所述待识别用户的语音数据输入至调整后的预设语音识别模型进行语音识别，确定
 所述待识别用户对应的语音识别结果。
- 40 8、一种语音数据生成装置，其中，包括：
 获取单元，用于获取目标用户的样本语音数据；
 提取单元，用于对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语
 音特征；
 第一确定单元，用于根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，

计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过对所述样本语音数据进行训练得到的;

第二确定单元,用于基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据。

5 9、一种计算机可读存储介质,其上存储有计算机程序,其中,所述计算机程序被处理器执行时实现语音数据生成方法;

其中,所述语音数据生成方法的步骤包括:

获取目标用户的样本语音数据;

对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征;

10 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,所述嵌入矩阵是通过对所述样本语音数据进行训练得到的;

基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据。

10、根据权利要求9所述的计算机可读存储介质,其中,所述根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数据对应的注意力分值,包括:

15 根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量;

将所述语音特征对应的查询向量和键向量相乘,得到所述语音特征对应的权重值;

根据所述语音特征对应的权重值和值向量,计算所述语音特征对应的注意力分值。

11、根据权利要求9所述的计算机可读存储介质,其中,所述对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征,包括:

20 对所述样本语音数据进行滤波处理,得到所述样本语音数据对应的语音能量;

对所述语音能量进行离散余弦化处理,得到所述样本语音数据对应的语音特征。

12、根据权利要求9所述的计算机可读存储介质,其中,在所述基于所述注意力分值,确定所述目标用户样本语音数据之外的验证语音数据之后,所述方法还包括:

25 将所述目标用户的样本语音数据和验证语音数据,以及预设样本库中其他用户的样本语音数据确定为第一训练样本;

利用预设神经网络算法对所述第一训练样本进行训练,构建预设语音识别模型。

13、根据权利要求12所述的计算机可读存储介质,其中,在所述利用预设神经网络算法对所述第一训练样本进行训练,构建预设语音识别模型之后,所述方法还包括:

30 将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本;

利用所述第二训练样本对所述预设语音识别模型进行调整,得到调整后的预设语音识别模型。

14、根据权利要求13所述的计算机可读存储介质,其中,所述方法还包括:

35 利用测试样本对所述调整后的预设语音识别模型进行测试,得到所述调整后的预设语音识别模型对应的测试结果;

根据所述测试结果,确定所述调整后的预设语音识别模型对应的语音识别准确率;

若所述语音识别准确率小于预设语音识别准确率,对所述调整后的预设语音识别模型中的参数进行调整,直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。

40 15、一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其中,所述计算机程序被处理器执行时实现语音数据生成方法;

其中,所述语音数据生成方法的步骤包括:

获取目标用户的样本语音数据;

对所述样本语音数据进行特征提取,得到所述样本语音数据对应的语音特征;

45 根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵,计算所述样本语音数

据对应的注意力分值，所述嵌入矩阵是通过对所述样本语音数据进行训练得到的；

基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据。

16、根据权利要求15所述的计算机设备，其中，所述根据所述样本语音数据对应的语音特征和预先构建的嵌入矩阵，计算所述样本语音数据对应的注意力分值，包括：

5 根据所述嵌入矩阵确定所述语音特征对应的查询向量、键向量和值向量；

将所述语音特征对应的查询向量和键向量相乘，得到所述语音特征对应的权重值；

根据所述语音特征对应的权重值和值向量，计算所述语音特征对应的注意力分值。

17、根据权利要求15所述的计算机设备，其中，所述对所述样本语音数据进行特征提取，得到所述样本语音数据对应的语音特征，包括：

10 对所述样本语音数据进行滤波处理，得到所述样本语音数据对应的语音能量；

对所述语音能量进行离散余弦化处理，得到所述样本语音数据对应的语音特征。

18、根据权利要求15所述的计算机设备，其中，在所述基于所述注意力分值，确定所述目标用户样本语音数据之外的验证语音数据之后，所述方法还包括：

15 将所述目标用户的样本语音数据和验证语音数据，以及预设样本库中其他用户的样本语音数据确定为第一训练样本；

利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型。

19、根据权利要求18所述的计算机设备，其中，在所述利用预设神经网络算法对所述第一训练样本进行训练，构建预设语音识别模型之后，所述方法还包括：

20 将所述目标用户的样本语音数据和所述其他用户的样本语音数据确定为第二训练样本；

利用所述第二训练样本对所述预设语音识别模型进行调整，得到调整后的预设语音识别模型。

20、根据权利要求19所述的计算机设备，其中，所述方法还包括：

25 利用测试样本对所述调整后的预设语音识别模型进行测试，得到所述调整后的预设语音识别模型对应的测试结果；

根据所述测试结果，确定所述调整后的预设语音识别模型对应的语音识别准确率；

若所述语音识别准确率小于预设语音识别准确率，对所述调整后的预设语音识别模型中的参数进行调整，直至所述调整后的预设语音识别模型对应的语音识别准确率达到预设语音识别准确率。

30

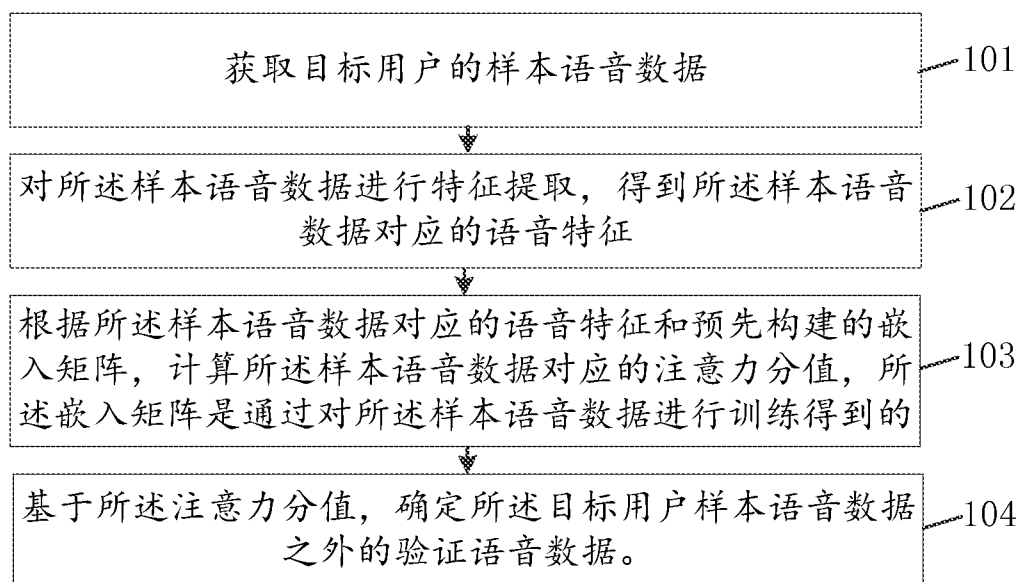


图 1

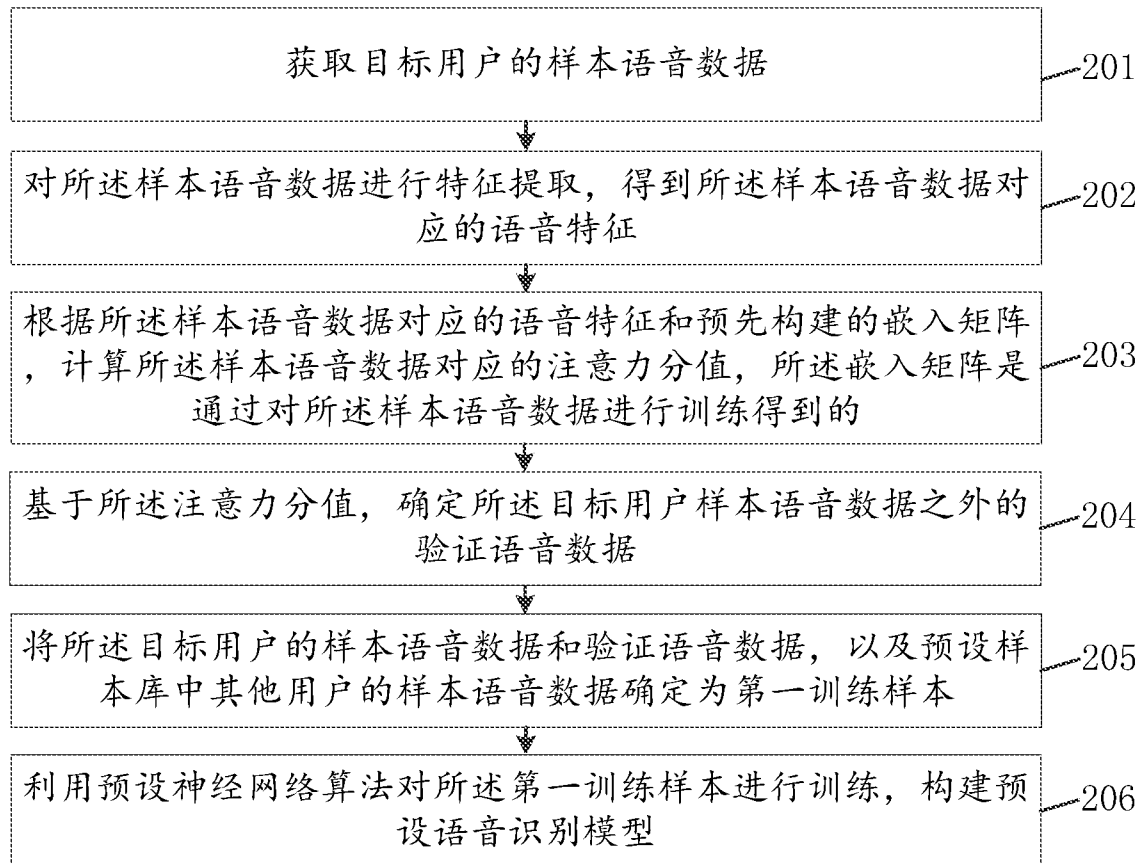


图 2

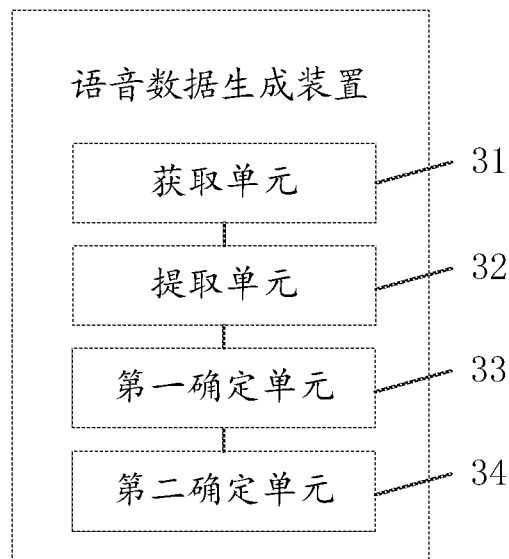


图 3

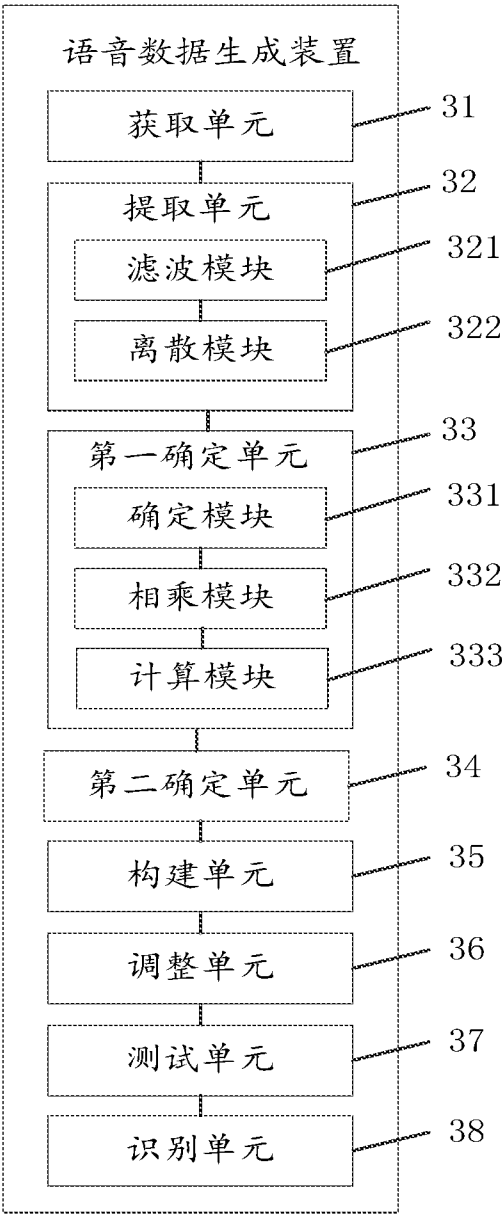


图 4

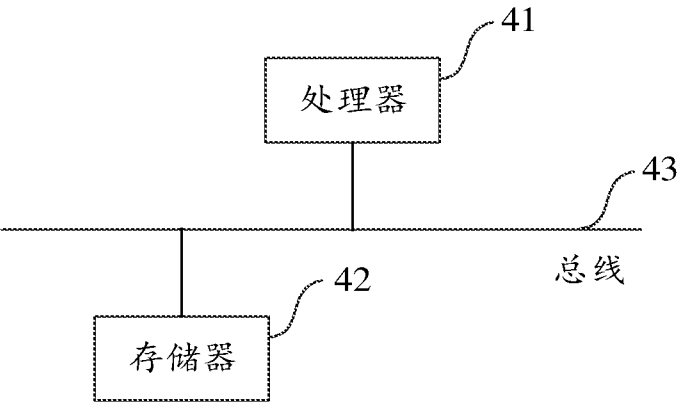


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/136366

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/16(2006.01)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)

G10L

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, IEEE: 语音, 注意力, 扩充, 扩展, 增加, 数据集, 样本, 特征, 选择, 增强, 计算, audio, attention, augment+, extend, sample, select+, feature, compute

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 111798874 A (NORTHWEST NORMAL UNIVERSITY) 20 October 2020 (2020-10-20) description paragraphs 0075-145	1-20
Y	胡婷婷 等 (HU, Tingting et al.). "基于注意力机制的LSTM语音情感主要特征选择 (The Salient Feature Selection by Attention Mechanism Based LSTM in Speech Emotion Recognition)" 声学技术 (Technical Acoustics), Vol. 38, No. 4, 31 August 2019 (2019-08-31), full text, pages 416-419	1-20
A	CN 110992938 A (TONGDUN HOLDINGS CO., LTD.) 10 April 2020 (2020-04-10) entire document	1-20
A	CN 111429938 A (JIANGSU UNIVERSITY) 17 July 2020 (2020-07-17) entire document	1-20
A	US 2019189112 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD.) 20 June 2019 (2019-06-20) entire document	1-20
Y	CN 109817246 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 28 May 2019 (2019-05-28) description paragraphs 0057-0119	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

11 July 2021

Date of mailing of the international search report

30 July 2021

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/CN2020/136366

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111798874	A	20 October 2020	None			
CN	110992938	A	10 April 2020	None			
CN	111429938	A	17 July 2020	None			
US	2019189112	A1	20 June 2019	CN	106251859	B	31 May 2019
				CN	106251859	A	21 December 2016
				JP	6774551	B2	28 October 2020
				WO	2018014469	A1	25 January 2018
				JP	2019527852	A	03 October 2019
CN	109817246	A	28 May 2019	WO	2020173133	A1	03 September 2020

国际检索报告

国际申请号

PCT/CN2020/136366

A. 主题的分类 G10L 15/16 (2006.01) 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) G10L 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, IEEE: 语音, 注意力, 扩充, 扩展, 增加, 数据集, 样本, 特征, 选择, 增强, 计算, audio, attention, augment+, extend, sample, select+, feature, compute																							
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>Y</td> <td>CN 111798874 A (西北师范大学) 2020年 10月 20日 (2020 - 10 - 20) 说明书第0075-145段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>胡婷婷 等. “基于注意力机制的LSTM语音情感主要特征选择” 声学技术, 第38卷, 第4期, 2019年 8月 31日 (2019 - 08 - 31), 正文第416-419页</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 110992938 A (同盾控股有限公司) 2020年 4月 10日 (2020 - 04 - 10) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 111429938 A (江苏大学) 2020年 7月 17日 (2020 - 07 - 17) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2019189112 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD.) 2019 年 6月 20日 (2019 - 06 - 20) 全文</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 109817246 A (平安科技深圳有限公司) 2019年 5月 28日 (2019 - 05 - 28) 说明书第0057-0119段</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	Y	CN 111798874 A (西北师范大学) 2020年 10月 20日 (2020 - 10 - 20) 说明书第0075-145段	1-20	Y	胡婷婷 等. “基于注意力机制的LSTM语音情感主要特征选择” 声学技术, 第38卷, 第4期, 2019年 8月 31日 (2019 - 08 - 31), 正文第416-419页	1-20	A	CN 110992938 A (同盾控股有限公司) 2020年 4月 10日 (2020 - 04 - 10) 全文	1-20	A	CN 111429938 A (江苏大学) 2020年 7月 17日 (2020 - 07 - 17) 全文	1-20	A	US 2019189112 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD.) 2019 年 6月 20日 (2019 - 06 - 20) 全文	1-20	Y	CN 109817246 A (平安科技深圳有限公司) 2019年 5月 28日 (2019 - 05 - 28) 说明书第0057-0119段	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
Y	CN 111798874 A (西北师范大学) 2020年 10月 20日 (2020 - 10 - 20) 说明书第0075-145段	1-20																					
Y	胡婷婷 等. “基于注意力机制的LSTM语音情感主要特征选择” 声学技术, 第38卷, 第4期, 2019年 8月 31日 (2019 - 08 - 31), 正文第416-419页	1-20																					
A	CN 110992938 A (同盾控股有限公司) 2020年 4月 10日 (2020 - 04 - 10) 全文	1-20																					
A	CN 111429938 A (江苏大学) 2020年 7月 17日 (2020 - 07 - 17) 全文	1-20																					
A	US 2019189112 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD.) 2019 年 6月 20日 (2019 - 06 - 20) 全文	1-20																					
Y	CN 109817246 A (平安科技深圳有限公司) 2019年 5月 28日 (2019 - 05 - 28) 说明书第0057-0119段	1-20																					
☐ 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 style="vertical-align: top;"> * 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 style="vertical-align: top;"> “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 2021年 7月 11日		国际检索报告邮寄日期 2021年 7月 30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 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/CN2020/136366

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	111798874	A	2020年 10月 20日	无			
CN	110992938	A	2020年 4月 10日	无			
CN	111429938	A	2020年 7月 17日	无			
US	2019189112	A1	2019年 6月 20日	CN	106251859	B	2019年 5月 31日
				CN	106251859	A	2016年 12月 21日
				JP	6774551	B2	2020年 10月 28日
				WO	2018014469	A1	2018年 1月 25日
				JP	2019527852	A	2019年 10月 3日
CN	109817246	A	2019年 5月 28日	WO	2020173133	A1	2020年 9月 3日