

(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
2022 年 12 月 29 日 (29.12.2022)



(10) 国际公布号
WO 2022/267451 A1

(51) 国际专利分类号:
G10L 15/32 (2013.01) *G06N 3/02* (2006.01)
G10L 15/06 (2013.01) *G06K 9/62* (2022.01)
G10L 15/14 (2006.01) *G06F 40/284* (2020.01)
G10L 15/16 (2006.01)

(21) 国际申请号: PCT/CN2022/071220

(22) 国际申请日: 2022 年 1 月 11 日 (11.01.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110706592.9 2021年6月24日 (24.06.2021) 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 518033 (CN)。

(72) 发明人: 方明(FANG, Ming); 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 518033 (CN)。魏韬(WEI, Tao); 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 518033 (CN)。马骏(MA, Jun); 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 518033 (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 518033 (CN)。

(74) 代理人: 北京鸿元知识产权代理有限公司(GRANDER IP LAW FIRM); 中国北京市朝阳区光华路 7 号汉威大厦东区 18A6 室, Beijing 100004 (CN)。

(54) Title: AUTOMATIC SPEECH RECOGNITION METHOD BASED ON NEURAL NETWORK, DEVICE, AND READABLE STORAGE MEDIUM

(54) 发明名称: 基于神经网络的自动语音识别方法、设备及可读存储介质

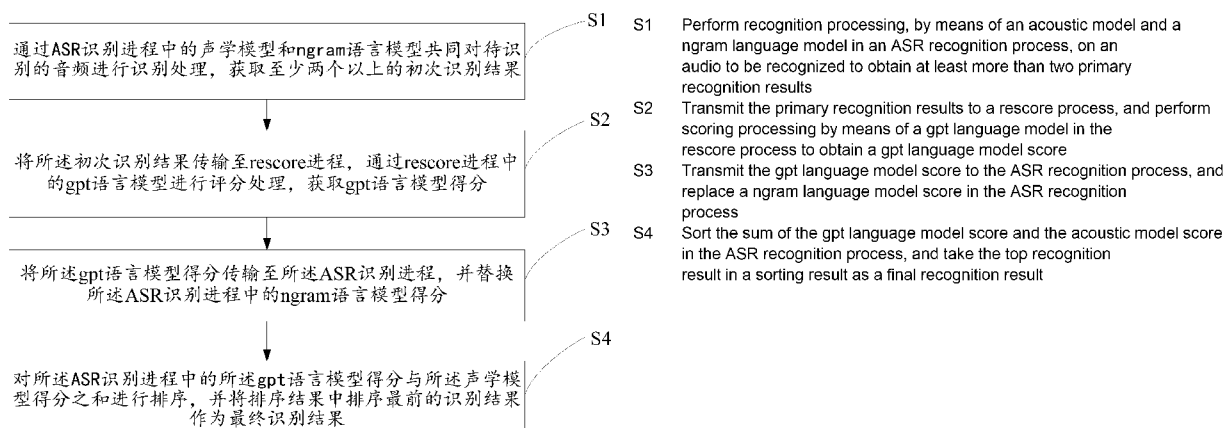


图 1

(57) Abstract: The present application relates to artificial intelligence, and provides an automatic speech recognition method and apparatus based on a neural network, an electronic device, and a computer-readable storage medium. The method comprises: performing recognition processing, by means of an acoustic model and a ngram language model in an ASR recognition process, on an audio to be recognized to obtain at least more than two primary recognition results; transmitting the primary recognition results to a rescore process, and performing scoring processing by means of a gpt language model in the rescore process to obtain a gpt language model score; transmitting the gpt language model score to the ASR recognition process, and replacing a ngram language model score in the ASR recognition process; and sorting the sum of the gpt language model score and the acoustic model score in the ASR recognition process, and taking the top recognition result in a sorting result as a final recognition result. The present application mainly aims to solve the problem of data sparsity by using the gpt language model.

(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))。

(57) 摘要: 本申请涉及一种人工智能, 提供一种基于神经网络的自动语音识别方法、装置、电子设备及计算机可读存储介质, 其中方法包括: 通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理, 获取至少两个以上的初次识别结果; 将所述初次识别结果传输至rescore进程, 并通过rescore进程中的gpt语言模型进行评分处理, 获取gpt语言模型得分; 将所述gpt语言模型得分传输至所述ASR识别进程, 并替换所述ASR识别进程中的ngram语言模型得分; 对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序, 并将排序结果中排序最前的识别结果作为最终识别结果。本申请主要目的在于通过采用gpt语言模型, 解决数据稀疏性的问题。

基于神经网络的自动语音识别方法、设备及可读存储介质

[0001] 本申请要求于2021年06月24日提交中国专利局、申请号为202110706592.9，发明名称为“基于神经网络的自动语音识别方法、设备及可读存储介质”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。

技术领域

[0002] 本申请涉及人工智能技术领域，尤其涉及一种基于神经网络的自动语音识别方法、装置、电子设备及计算机可读存储介质。

背景技术

[0003] 在传统的语音识别的过程中包括两个模型，分别是声学模型和语言模型；其中，语言模型一般采用ngram语言模型，ngram这种基于元组统计的概率模型，只能抓到词组前后的统计信息，无法学习到更深入的语法、语义信息，再加上这种词频统计的概率计算方法，有参数空间过大问题和数据稀疏严重的问题，尤其在高阶ngram模型中，随着阶数增加，ngram模型和稀疏性会指数级别增加。即使人们提出和很多种办法去努力解决ngram模型本身的问题，例如剪枝和回退，都只是减弱ngram模型的问题，无法从解决ngram语言模型的根本性问题。

[0004] 目前一种常见的解决方案：保持原有ngram模型不变，在wfst解码之后，生成top n的ARS识别结果的基础上，重新对生成语句的语言模型进行打分，并进行重排序。发明人意识到，采用有更多语料的ngram模型，更高阶的ngram模型等等；但常常遇到的问题是采用越复杂语言模型常常会导致更多的识别时延，用简单的语言模型往往不能取得准确的识别效果。

[0005] 为了解决上述问题，亟需一种新的自动语音识别方案。

发明概述

技术问题

[0006] 本申请提供一种基于神经网络的自动语音识别方法、装置、电子设备及计算机可读存储介质，其主要目的在于通过采用gpt语言模型，解决数据稀疏性的问题

- 。
- [0007] 为实现上述目的，本申请提供的基于神经网络的自动语音识别方法，应用于电子设备，所述方法包括：
- [0008] 通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；
- [0009] 将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；
- [0010] 将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；
- [0011] 对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。
- [0012] 为了解决上述问题，本申请还提供一种基于神经网络的自动语音识别装置，包括：
- [0013] 初次识别结果获取模块，用于通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；
- [0014] gpt语言模型得分获取模块，用于将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；
- [0015] 语言模型得分替换模块，用于将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；
- [0016] 最终识别结果获取模块，用于对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。
- [0017] 为了解决上述问题，本申请还提供一种电子设备，所述电子设备包括：
- [0018] 至少一个处理器；以及，

[0019] 与所述至少一个处理器通信连接的存储器；其中，

[0020] 所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行上述的基于神经网络的自动语音识别方法的步骤。

[0021] 为了解决上述问题，本申请还提供一种计算机可读存储介质，所述计算机可读存储介质中存储有至少一个指令，所述至少一个指令被电子设备中的处理器执行以实现上述所述的基于神经网络的自动语音识别方法。

[0022] 本申请实施例通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果述所述gpt语言模型得分与所述声学模型得分的排序的结果作为最终识别结果。本申请的主要目的在于通过采用gpt语言模型，解决数据稀疏性的问题。

问题的解决方案

技术解决方案

[0023] 在此处键入技术解决方案描述段落。

发明的有益效果

有益效果

[0024] 在此处键入有益效果描述段落。

对附图的简要说明

附图说明

[0025] 图1为本申请一实施例提供的基于神经网络的自动语音识别方法的流程示意图；

[0026] 图2为本申请一实施例提供的基于神经网络的自动语音识别装置的模块示意图；

[0027] 图3为本申请一实施例提供的实现基于神经网络的自动语音识别方法的电子设备的内部结构示意图；

[0028] 本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

实施该发明的最佳实施例

本发明的最佳实施方式

[0029] 在此处键入本发明的最佳实施方式描述段落。

发明实施例

本发明的实施方式

[0030] 为解决上述问题，本申请提供一种基于神经网络的自动语音识别方法。参照图1所示，为本申请一实施例提供的基于神经网络的自动语音识别方法的流程示意图。该方法可以由一个装置执行，该装置可以由软件和/或硬件实现。

[0031] 在本实施例中，基于神经网络的自动语音识别方法包括：

[0032] S1：通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；

[0033] S2：将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；

[0034] S3：将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；

[0035] S4：对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。

[0036] 上述为本申请人工智能的基于神经网络的自动语音识别方法，在本申请的基于神经网络的自动语音识别方法中，包含ASR（自动语音识别，Automatic Speech Recognition，简称“ASR”）识别进程和rescore（重打分）进程。ASR识别进程为音转字模块，采用传统的GMM-HMM技术路线，神经网络的建模单元为hmm state，输入为声学特征，经历声学模型和语言模型解码两个模块完成ASR识别。在识别开始前，要进行声学模型与语言模型的训练，声学模型的训练原理为

神经网络的反向传播，需要大量的音频时长与标注文本；语言模型训练输入为大量的文本语料，做3gram语言模型建模生成arpa格式的语言模型，接着利用arpa2fst等工具生成hclg结构的wfst图，并作为输入用于语言模型解码。

[0037] 在步骤S1中，所述通过ASR识别进程中声学模型和ngram语言模型对待识别的音频进行识别处理，获取至少两个以上的识别结果，包括如下步骤：

[0038] S111：将所述待识别的音频转化为音频特征；

[0039] S112：根据所述音频特征获取所述音频特征中每帧的后验概率；

[0040] S113：根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码，生成lattice图；以及

[0041] S114：根据所述lattice图，获取top n个识别结果；其中，每个识别结果包含文本、声学模型得分、ngram语言模型得分及两个模型（声学模型、ngram语言模型）的得分之和。

[0042] 其中，在步骤S111中，将所述待识别的音频转化为音频特征，包括如下步骤：

[0043] 步骤S11101：对所述待识别音频语音进行分帧、加窗处理，获取规范音频；以及，

[0044] 步骤S11102：通过MFCC特征提取算法对所述规范音频进行特征提取，获取所述待识别音频的音频特征。

[0045] 在步骤S112中，所述将根据所述音频特征获取所述音频特征的中每帧的后验概率，包括如下步骤：

[0046] 步骤S11201：将所述音频特征提取为音频特征向量序列；

[0047] 步骤S11202：将所述音频特征向量序列音频特征 输入预先训练好的声学模型，确定音素状态的时间边界；

[0048] 步骤S11203：根据所述时间边界，提取所述时间边界内的所有帧，按语音帧的帧长取平均值，作为所述语音帧的后验概率。

[0049] 其中，在步骤S113中，所述根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码，生成lattice图，包括如下步骤：

[0050] 步骤S11301：将ngram语言模型建模生成arpa格式的语言模型；

[0051] S11302：利用arpa2fst工具生成hclg结构的wfst图；

- [0052] S11303: 根据维特比算法 (viterbi)、所述后验概率及所述wfst图构建wfst搜索空间;
- [0053] S11304: 在加权有限状态转换器(weightedfinite-statetransducer, wfst)搜索空间内寻找匹配概率最大的最优路径, 得到文字识别结果。
- [0054] 其中, 每个识别结果包含文本、声学模型得分、语言模型得分及声学模型得分和语言模型得分之和。
- [0055] 在本申请的实施例中, lattice图中可取出声学模型和语言模型得分。对lattice的每个输出做总得分的从小到大排序, 回溯top1结果即为ngram语言模型输出的默认ASR结果, 回溯top n结果则可以提取nbest信息, 输出到rescore进程完成重打分工作。
- [0056] 在步骤S120中, rescore线程为重打分模块, 因有gpu依赖被单独设计为一个进程, 工作在gpu上面并使用TensorRT加速, 除完成gpt语言模型的推理过程, 还要负责解码线程的请求与响应。Rescore模块每次会输入固定batch大小的文本语句, 输出每条文本对应的gpt语言模型得分。
- [0057] 其中, 所述将所述识别结果传输至rescore进程, 并通过rescore进程中的gpt语言模型进行处理, 获取gpt语言模型得分, 包括如下步骤:
- [0058] 步骤S121: 在预设时间内, 将待rescore (重打分) 的语句拼凑成批量的待重新打分语句;
- [0059] 步骤S122: 通过gpt语言模型对批量的待重新打分语句做神经网络前向推理;
- [0060] 步骤S123: 累加所述待重新打分语句的每个词的后验概率, 并以对数形式输出, 以获取待重新打分语句的gpt语言模型得分。
- [0061] 在本申请的一个具体的实施例中, 例假设要打分的文本为 “[CLS] the dog is hairy [SEP]”, 那么输入到gpt模型文本序列为 “[CLS] the dog is hairy” 共5个词, 在输出概率矩阵上取当前词对应下个词的概率, 具体到本例中, “the” 这个词的输出概率分布中, 当前词的下个词为 “dog”, 则取 “the” 的概率分布项中 “dog” 词的概率项为其对应的输出概率, 经此处理后, 假定上述输入序列的输出对数概率序列为 p_1 p_2 p_3 p_4 p_5 , 累加 p_1 到 p_5 即为gpt语言模型得分。

[0062] 在步骤S130和步骤S140中，在rescore进程再将gpt语言模型得分结果返回到ASR解码线程中，ASR解码线程用gpt语言模型得分替换掉top n语句中总得分中的ngram得分，再对新的总得分进行从小到大重排序，使用重排序的top 1的ASR文本作为最终的ASR识别结果，也就是说，将gpt语言模型得分、声学模型得分排序结果中排序最前的识别结果作为最终识别结果，从而达到提升ASR准确率的目的。

[0063] 在本申请的实施例，对ASR识别结果做GPT rescore，经过实验测试，对asr整体的识别结果的准确率有较高的提升，往往可以提升2个百分点左右，字准下降1个百分点左右。从时延的角度，仅会造成50ms左右的时延，对于整体语音识别系统的延时影响是很有限的。而识别准确率的提升，不仅仅是单纯的字准的下降，带来更好的ASR体验。还有对于依赖ASR识别结果的上游系统，例如语音客服机器人、智能语音助手、智能音箱等，都可以间接提升上游系统的效果，提升服务质量，提高客户满意度。

[0064] 本申请实施例通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。本申请的主要目的在于通过采用gpt语言模型，解决数据稀疏性的问题。

[0065] 如图2所示，是本申请基于神经网络的自动语音识别装置的功能模块图。

[0066] 本申请所述基100可以安装于电子设备中。根据实现的功能，所述基于神经网络的自动语音识别装置可以包括：初次识别结果获取模块101、gpt语言模型得分获取模块102、语言模型得分替换模块103和最终识别结果获取模块104。本申请所述模块也可以称之为单元，是指一种能够被电子设备处理器所执行，并且能够完成固定功能的一系列计算机程序段，其存储在电子设备的存储器中。

[0067] 在本实施例中，关于各模块/单元的功能如下：

- [0068] 初次识别结果获取模块101，用于通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；
- [0069] gpt语言模型得分获取模块102，用于将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；
- [0070] 语言模型得分替换模块103，用于将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；
- [0071] 最终识别结果获取模块104，用于对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。
- [0072] 在本申请的实施例中，初次识别结果获取模块101包括音频特征转化模块、后验概率获取模块、lattice图获取模块和两个以上的识别结果获取模块，其中，
- [0073] 所述音频特征转化模块，用于将所述待识别的音频转化为音频特征；
- [0074] 所述后验概率获取模块，用于根据所述音频特征获取所述音频特征中每帧的后验概率；
- [0075] 所述lattice图获取模块，用于根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图；
- [0076] 所述两个以上的识别结果获取模块，用于根据所述lattice图，获取至少两个以上的初次识别结果。
- [0077] 其中，音频特征转化模块包括规范音频获取模块和待识别音频的音频特征获取模块。其中，
- [0078] 所述规范音频获取模块，用于对所述待识别音频语音进行分帧、加窗处理，获取规范音频；以及，
- [0079] 所述待识别音频的音频特征获取模块，用于通过MFCC特征提取算法对所述规范音频进行特征提取，获取所述待识别音频的音频特征。
- [0080] 其中，所述后验概率获取模块，用于将根据所述音频特征获取所述音频特征的中每帧的后验概率，包括：音频特征向量序获取模块、音素状态的时间边界确

定模块、后验概率确定模块。其中，

[0081] 音频特征向量序列获取模块，用于将所述音频特征提取为音频特征向量序列；

[0082] 音素状态的时间边界确定模块，用于将所述音频特征向量序列音频特征 输入预先训练好的声学模型，确定音素状态的时间边界；

[0083] 后验概率确定模块，用于根据所述时间边界，提取所述时间边界内的所有帧，按语音帧的帧长取平均值，作为所述语音帧的后验概率。

[0084] 其中，lattice图生成模块，包括：arpa格式生成模块、wfst图生成模块、wfst搜索空间构建模块和文字识别结果确定模块，其中，

[0085] arpa格式生成模块，用于将ngram语言模型建模生成arpa格式的语言模型；

[0086] wfst图生成模块，用于利用arpa2fst工具生成hclg结构的wfst图；

[0087] wfst搜索空间构建模块，用于根据维特比算法（viterbi）、所述后验概率及所述wfst图构建wfst搜索空间；

[0088] 文字识别结果确定模块，用于在所述wfst搜索空间搜索空间内寻找匹配概率最大的最优路径，得到文字识别结果。

[0089] 其中，wfst（weightedfinite-state transducer），具体指加权有限状态转换器。

[0090] 在本申请的实施例，每个识别结果包含文本、声学模型得分、语言模型得分及两个模型（声学模型、语言模型）的得分之和。

[0091] 在本申请的实施例，lattice图中可取出声学模型和语言模型得分。对lattice的每个输出做总得分的从小到大排序，回溯top1结果即为ngram语言模型输出的默认ASR结果，回溯top n结果则可以提取nbest信息，输出到rescore进程完成重打分工作。

[0092] 在所述gpt语言模型得分获取模块102中，rescore线程即为重打分模块，因有gpu依赖被单独设计为一个进程，工作在gpu上面并使用TensorRT加速，除完成gpt语言模型的推理过程，还要负责解码线程的请求与响应。Rescore模块每次会输入固定batch大小的文本语句，输出每条文本对应的gpt语言模型得分。

[0093] 其中，在gpt语言模型得分获取模块102中，在预设时间内，将待rescore（重打分）的语句拼凑成批量的待重新打分语句；

- [0094] 通过gpt语言模型对批量的待重新打分语句做神经网络前向推理；
- [0095] 累加所述待重新打分语句的每个词的后验概率，并以对数形式输出，以获取待重新打分语句的gpt语言模型得分。
- [0096] 在本申请的一个具体的实施例中，例假设要打分的文本为“ [CLS] the dog is hairy [SEP]”，那么输入到gpt模型文本序列为“[CLS] the dog is hairy”共5个词，在输出概率矩阵上取当前词对应下个词的概率，具体到本例中，“the”这个词的输出概率分布中，当前词的下个词为“dog”，则取“the”的概率分布项中“dog”词的概率项为其对应的输出概率，经此处理后，假定上述输入序列的输出对数概率序列为 p_1 、 p_2 、 p_3 、 p_4 、 p_5 ，累加 p_1 到 p_5 即为：gpt语言模型得分。
- [0097] 在语言模型得分替换模块103和最终识别结果获取模块104中，rescore进程再将gpt语言模型得分结果返回到ASR解码线程，ASR解码线程用gpt语言模型得分替换掉top n语句中总得分中的ngram得分，再对新的总得分进行从小到大重排序，使用重排序的top 1的ASR文本作为最终的ASR识别结果（排序结果中排序最前的识别结果作为最终识别结果），从而达到提升ASR准确率的目的。
- [0098] 在本申请的实施例中，对ASR识别结果做GPT rescore，经过实验测试，对ASR整体的识别结果的准确率有较高的提升，往往可以提升2个百分点左右，字准下降1个百分点左右。从时延的角度，仅会造成50ms左右的时延，对于整体语音识别系统的延时影响是很有限的。而识别准确率的提升，不仅仅是单纯的字准的下降，带来更好的ASR体验。还有对于依赖ASR识别结果的上游系统，例如语音客服机器人、智能语音助手、智能音箱等，都可以间接提升上游系统的效果，提升服务质量，提高客户满意度。
- [0099] 在本申请的实施例中通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的

识别结果作为最终识别结果。本申请的主要目的在于通过采用gpt语言模型，解决数据稀疏性的问题。

[0100] 如图3所示，是本申请实现基于神经网络的自动语音识别方法的电子设备的结构示意图。

[0101] 所述电子设备1可以包括处理器10、存储器11和总线，还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序，如基于神经网络的自动语音识别程序12。

[0102] 其中，所述存储器11至少包括一种类型的可读存储介质，所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器（例如：SD或DX存储器等）、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备1的内部存储单元，例如该电子设备1的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备1的外部存储设备，例如电子设备1上配备的插接式移动硬盘、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，所述存储器11还可以既包括电子设备1的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备1的应用软件及各类数据，例如数据稽核程序的代码等，还可以用于暂时地存储已经输出或者将要输出的数据。存储器可以存储内容，该内容可由电子设备显示或被发送到其他设备（例如，耳机）以由其他设备来显示或播放。存储器还可以存储从其他设备接收的内容。该来自其他设备的内容可由电子设备显示、播放、或使用，以执行任何必要的可由电子设备和/或无线接入点中的计算机处理器或其他组件实现的任务或操作。

[0103] 所述处理器10在一些实施例中可以由集成电路组成，例如可以由单个封装的集成电路所组成，也可以是由多个相同功能或不同功能封装的集成电路所组成，包括一个或者多个中央处理器（Central Processing unit， CPU）、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心（Control Unit），利用各种接口和线路连接整个电子设备的各个部件，通过运行或执行存储在所述存储器11内的程序或者模块（例如数据稽核程序等），以及调用存储在所述存储器11内的数据，以执行电子设备1

的各种功能和处理数据。电子还可包括芯片组（未示出），其用于控制一个或多个处理器与用户设备的其他组件中的一个或多个之间的通信。在特定的实施例中，电子设备可基于Intel®架构或ARM®架构，并且处理器和芯片集可来自Intel®处理器和芯片集家族。该一个或多个处理器104还可包括一个或多个专用集成电路（ASIC）或专用标准产品（ASSP），其用于处理特定的数据处理功能或任务。

[0104] 所述总线可以是外设部件互连标准（peripheral component interconnect，简称PCI）总线或扩展工业标准结构（extended industry standard architecture，简称EISA）总线等。该总线可以分为地址总线、数据总线、控制总线等。所述总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。

[0105] 此外，网络和I/O接口可包括一个或多个通信接口或网络接口设备，以提供经由网络（未示出）在电子设备和其他设备（例如，网络服务器）之间的数据传输。通信接口可包括但不限于：人体区域网络（BAN）、个人区域网络（PAN）、有线局域网（LAN）、无线局域网（WLAN）、无线广域网（WWAN）、等等。用户设备102可以经由有线连接耦合到网络。然而，无线系统接口可包括硬件或软件以广播和接收消息，其使用Wi-Fi直连标准和/或IEEE 802.11无线标准、蓝牙标准、蓝牙低功耗标准、Wi-Gig标准、和/或任何其他无线标准和/或它们的组合。

[0106] 无线系统可包括发射器和接收器或能够在由IEEE 802.11无线标准所支配的操作频率的广泛范围内操作的收发器。通信接口可以利用声波、射频、光学、或其他信号来在电子设备与其他设备（诸如接入点、主机、服务器、路由器、读取设备、和类似物）之间交换数据。网络118可包括但不限于：因特网、专用网络、虚拟专用网络、无线广域网、局域网、城域网、电话网络、等等。

[0107] 显示器可包括但不限于液晶显示器、发光二极管显示器、或由在美国马萨诸塞州剑桥城的E Ink公司（E Ink Corp. of Cambridge, Massachusetts）所制造的E-Ink™显示器。该显示器可用于将内容以文本、图像、或视频的形式显示给用户。在特定的实例中，该显示器还可以作为触控屏显示器操作，其可以使得

用户能够藉由使用某些手指或手势来触摸屏幕以启动命令或操作。

[0108] 图3仅示出了具有部件的电子设备，本领域技术人员可以理解的是，图2示出的结构并不构成对所述电子设备1的限定，可以包括比图示更少或者更多的部件，或者组合某些部件，或者不同的部件布置。

[0109] 例如，尽管未示出，所述电子设备1还可以包括给各个部件供电的电源（比如电池），优选地，电源可以通过电源管理装置与所述至少一个处理器10逻辑相连，从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备1还可以包括多种传感器、蓝牙模块、Wi-Fi模块等，在此不再赘述。

[0110] 进一步地，所述电子设备1还可以包括网络接口，可选地，所述网络接口可以包括有线接口和/或无线接口（如WI-FI接口、蓝牙接口等），通常用于在该电子设备1与其他电子设备之间建立通信连接。

[0111] 可选地，该电子设备1还可以包括用户接口，用户接口可以是显示器（Display）、输入单元（比如键盘（Keyboard）），可选地，用户接口还可以是标准的有线接口、无线接口。可选地，在一些实施例中，显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED（Organic Light-Emitting Diode，有机发光二极管）触摸器等。其中，显示器也可以适当的称为显示屏或显示单元，用于显示在电子设备1中处理的信息以及用于显示可视化的用户界面。

[0112] 应该了解，所述实施例仅为说明之用，在专利申请范围上并不受此结构的限制。

[0113] 所述电子设备1中的所述存储器11存储的基于神经网络的自动语音识别程序12是多个指令的组合，在所述处理器10中运行时，可以实现：

[0114] 通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；

[0115] 将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模

型进行评分处理，获取gpt语言模型得分；

[0116] 将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；

[0117] 对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。

[0118] 具体地，所述处理器10对上述指令的具体实现方法可参考图1对应实施例中相关步骤的描述，在此不赘述。

[0119] 进一步地，所述电子设备1集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）。

[0120] 在本申请的实施例中，计算机可读存储介质，所述计算机可读存储介质中存储有至少一个指令，所述至少一个指令被电子设备中的处理器执行以实现上述所述的基于神经网络的自动语音识别方法的步骤，具体方法如下：

[0121] 通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；

[0122] 将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；

[0123] 将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；

[0124] 对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。

[0125] 在本申请所提供的几个实施例中，应该理解到，所揭露的设备，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，所述模块的划分，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。其中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。

[0126] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块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模块来实现本实施例方案的目的。

[0127] 另外，在本申请各个实施例中的各功能模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

[0128] 以上参考根据本申请的示例性实施例的系统和方法和/或计算机程序产品的框图和流程图描述了本申请的某些实施例。应当理解的是，框图和流程图中的一个或多个方框、以及在框图和流程图中的方框的组合，可以分别由计算机可执行程序指令实现。同样地，根据本申请的一些实施例，框图和流程图的一些方框可以不必按照所呈现的顺序执行，或者甚至可以完全不需要执行。

[0129] 这些计算机可执行程序指令可以被加载到通用计算机、专用计算机、处理器、或其他可编程数据处理装置上以产生特定机器，使得在计算机、处理器、或其他可编程数据处理装置上执行的指令创建用于实现在流程图方框或多个方框中所指定的一个或多个功能的构件。这些计算机程序产品还可以存储在计算机可读存储器中，其可以指导计算机或其他可编程数据处理装置以特定的方式运行，使得存储在计算机可读存储器中的指令产生制品，该制品包括实现在流程图的方框或多个方框中指定的一个或多个功能的指令构件。例如，本申请的实施例可提供计算机程序产品，其包括其中包含有计算机可读程序代码或程序指令的计算机可用介质，所述计算机可读程序代码适于被执行以实现在流程图方框或多个方框中指定的一个或多个功能。计算机程序指令还可以被加载到计算机或其他可编程数据处理装置上，以致使一系列操作元素或步骤在计算机或其他可编程装置上执行易产生计算机实现的程序，使得在计算机或其他可编程装置上执行的指令提供用于实现在流程图方框或多个方框中指定的功能的元素或步骤。

[0130] 相应地，框图或流程图中的方框支持用以执行指定功能的构件的组合、用于执行指定功能的元素或步骤与用于执行指定功能的程序指令构件的组合。还应当理解的是，框图和流程图中的每个方框以及框图和流程图中的方框的组合可由执行指定功能、元素或步骤的专用的基于硬件的计算机系统实现，或由专用硬件或计算机指令的组合实现。

[0131] 虽然本申请的某些实施例已经结合目前被认为是最实用的且各式各样的实施例进行了描述，但应当理解，本申请并不限于所公开的实施例，而是意在覆盖包含在所附权利要求书的范围之内的各种修改和等价布置。虽然本文采用了特定的术语，但它们仅以一般性和描述性的意义使用，而不是用于限制的目的。

[0132] 对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请的精神或基本特征的情况下，能够以其他的具体形式实现本申请。

[0133] 因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0134] 最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

工业实用性

[0135] 在此处键入工业实用性描述段落。

序列表自由内容

[0136] 在此处键入序列表自由内容描述段落。

权利要求书

- [权利要求 1] 一种基于神经网络的自动语音识别方法，应用于电子设备，其中，所述方法包括：
- 通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；
- 将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；
- 将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；
- 对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。
- [权利要求 2] 如权利要求1所述的基于神经网络的自动语音识别方法，其中，所述通过ASR识别进程中声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果，包括如下步骤：
- 将所述待识别的音频转化为音频特征；
- 根据所述音频特征获取所述音频特征中每帧的后验概率；
- 根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图；以及
- 根据所述lattice图，获取至少两个以上的初次识别结果。
- [权利要求 3] 如权利要求2所述的基于神经网络的自动语音识别方法，其中，所述将所述待识别的音频转化为音频特征，包括如下步骤：
- 对所述待识别音频进行分帧、加窗处理，获取规范音频；以及
- 通过MFCC特征提取算法对所述规范音频进行特征提取，获取所述待识别音频的音频特征。

- [权利要求 4] 如权利要求2所述的基于神经网络的自动语音识别方法，其中，所述根据所述音频特征获取所述音频特征中每帧的后验概率，包括如下步骤：
- 将所述音频特征提取为音频特征向量序列；
- 将所述音频特征向量序列输入预先训练好的声学模型，确定音素状态的时间边界；
- 根据所述时间边界，提取所述时间边界内的所有帧，按语音帧的帧长取平均值，作为所述语音帧的后验概率。
- [权利要求 5] 如权利要求2所述的基于神经网络的自动语音识别方法，其中，所述根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图，包括如下步骤：
- 将所述ngram语言模型建模生成arpa格式的语言模型；
- 利用arpa2fst工具生成hclg结构的wfst图；
- 根据所述viterbi、所述后验概率及所述wfst图构建wfst搜索空间；
- 在所述wfst搜索空间内寻找匹配概率最大的最优路径，得到文字识别结果。
- [权利要求 6] 如权利要求1所述的基于神经网络的自动语音识别方法，其中，所述将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行处理，获取gpt语言模型得分，包括如下步骤：
- 在预设时间内，将待重打分的语句拼凑成批量的待重新打分语句；
- 通过gpt语言模型对批量的待重新打分语句做神经网络前向推理；
- 累加所述待重新打分语句的每个词的后验概率，并以对数形式输出，以获取待重新打分语句的gpt语言模型得分。
- [权利要求 7] 一种基于神经网络的自动语音识别装置，其中，所述装置包括：
- 初次识别结果获取模块，用于通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；

gpt语言模型得分获取模块，用于将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；

语言模型得分替换模块，用于将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；

最终识别结果获取模块，用于对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。

[权利要求 8] 如权利要求7所述的基于神经网络的自动语音识别装置，其中，所述初次识别结果获取模块包括音频特征转化模块、后验概率获取模块、lattice图获取模块和两个以上的识别结果获取模块，其中，所述音频特征转化模块，用于将所述待识别的音频转化为音频特征；所述后验概率获取模块，用于根据所述音频特征获取所述音频特征中每帧的后验概率；所述lattice图获取模块，用于根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图；所述两个以上的识别结果获取模块，用于根据所述lattice图，获取至少两个以上的初次识别结果。

[权利要求 9] 一种电子设备，其中，所述电子设备包括：至少一个处理器；以及，与所述至少一个处理器通信连接的存储器；其中，所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行基于神经网络的自动语音识别方法的步骤，其中，所述基于神经网络的自动语音识别方法包括：通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与

所述ngram语言模型得分之和；

将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；

将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；

对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果。

[权利要求 10] 如权利要求9所述的电子设备，其中，
所述通过ASR识别进程中声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果，包括如下步骤：

将所述待识别的音频转化为音频特征；

根据所述音频特征获取所述音频特征中每帧的后验概率；

根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图；以及

根据所述lattice图，获取至少两个以上的初次识别结果。

[权利要求 11] 如权利要求10所述的电子设备，其中，所述将所述待识别的音频转化为音频特征，包括如下步骤：

对所述待识别音频进行分帧、加窗处理，获取规范音频；以及

通过MFCC特征提取算法对所述规范音频进行特征提取，获取所述待识别音频的音频特征。

[权利要求 12] 如权利要求10所述的电子设备，其中，所述根据所述音频特征获取所述音频特征中每帧的后验概率，包括如下步骤：

将所述音频特征提取为音频特征向量序列；

将所述音频特征向量序列输入预先训练好的声学模型，确定音素状态的时间边界；

根据所述时间边界，提取所述时间边界内的所有帧，按语音帧的帧长

取平均值，作为所述语音帧的后验概率。

[权利要求 13] 如权利要求10所述的电子设备，其中，
所述根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图，包括如下步骤：
将所述ngram语言模型建模生成arpa格式的语言模型；
利用arpa2fst工具生成hclg结构的wfst图；
根据所述viterbi、所述后验概率及所述wfst图构建wfst搜索空间；
在所述wfst搜索空间内寻找匹配概率最大的最优路径，得到文字识别结果。

[权利要求 14] 如权利要求9所述的电子设备，其中，
所述将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行处理，获取gpt语言模型得分，包括如下步骤：
在预设时间内，将待重打分的语句拼凑成批量的待重新打分语句；
通过gpt语言模型对批量的待重新打分语句做神经网络前向推理；
累加所述待重新打分语句的每个词的后验概率，并以对数形式输出，以获取待重新打分语句的gpt语言模型得分。

[权利要求 15] 一种计算机可读存储介质，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基于神经网络的自动语音识别方法，其中，
通过ASR识别进程中的声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果；其中，每个识别结果包括声学模型得分、ngram语言模型得分及所述声学模型得分与所述ngram语言模型得分之和；
将所述初次识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行评分处理，获取gpt语言模型得分；
将所述gpt语言模型得分传输至所述ASR识别进程，并替换所述ASR识别进程中的ngram语言模型得分；
对所述ASR识别进程中的所述gpt语言模型得分与所述声学模型得分之和进行排序，并将排序结果中排序最前的识别结果作为最终识别结果

。

- [权利要求 16] 如权利要求15所述的计算机可读存储介质，其中，
所述通过ASR识别进程中声学模型和ngram语言模型共同对待识别的音频进行识别处理，获取至少两个以上的初次识别结果，包括如下步骤：
将所述待识别的音频转化为音频特征；
根据所述音频特征获取所述音频特征中每帧的后验概率；
根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图；以及
根据所述lattice图，获取至少两个以上的初次识别结果。
- [权利要求 17] 如权利要求16所述的计算机可读存储介质，其中，所述将所述待识别的音频转化为音频特征，包括如下步骤：
对所述待识别音频进行分帧、加窗处理，获取规范音频；以及
通过MFCC特征提取算法对所述规范音频进行特征提取，获取所述待识别音频的音频特征。
- [权利要求 18] 如权利要求16所述的计算机可读存储介质，其中，所述根据所述音频特征获取所述音频特征中每帧的后验概率，包括如下步骤：
将所述音频特征提取为音频特征向量序列；
将所述音频特征向量序列输入预先训练好的声学模型，确定音素状态的时间边界；
根据所述时间边界，提取所述时间边界内的所有帧，按语音帧的帧长取平均值，作为所述语音帧的后验概率。
- [权利要求 19] 如权利要求16所述的计算机可读存储介质，其中，
所述根据所述每帧的后验概率，对ngram语言模型生成的wfst图进行viterbi解码生成lattice图，包括如下步骤：
将所述ngram语言模型建模生成arpa格式的语言模型；
利用arpa2fst工具生成hclg结构的wfst图；
根据所述viterbi、所述后验概率及所述wfst图构建wfst搜索空间；

在所述wfst搜索空间内寻找匹配概率最大的最优路径，得到文字识别结果。

[权利要求 20] 如权利要求15所述的计算机可读存储介质，其中，
所述将所述识别结果传输至rescore进程，并通过rescore进程中的gpt语言模型进行处理，获取gpt语言模型得分，包括如下步骤：
在预设时间内，将待重打分的语句拼凑成批量的待重新打分语句；
通过gpt语言模型对批量的待重新打分语句做神经网络前向推理；
累加所述待重新打分语句的每个词的后验概率，并以对数形式输出，
以获取待重新打分语句的gpt语言模型得分。

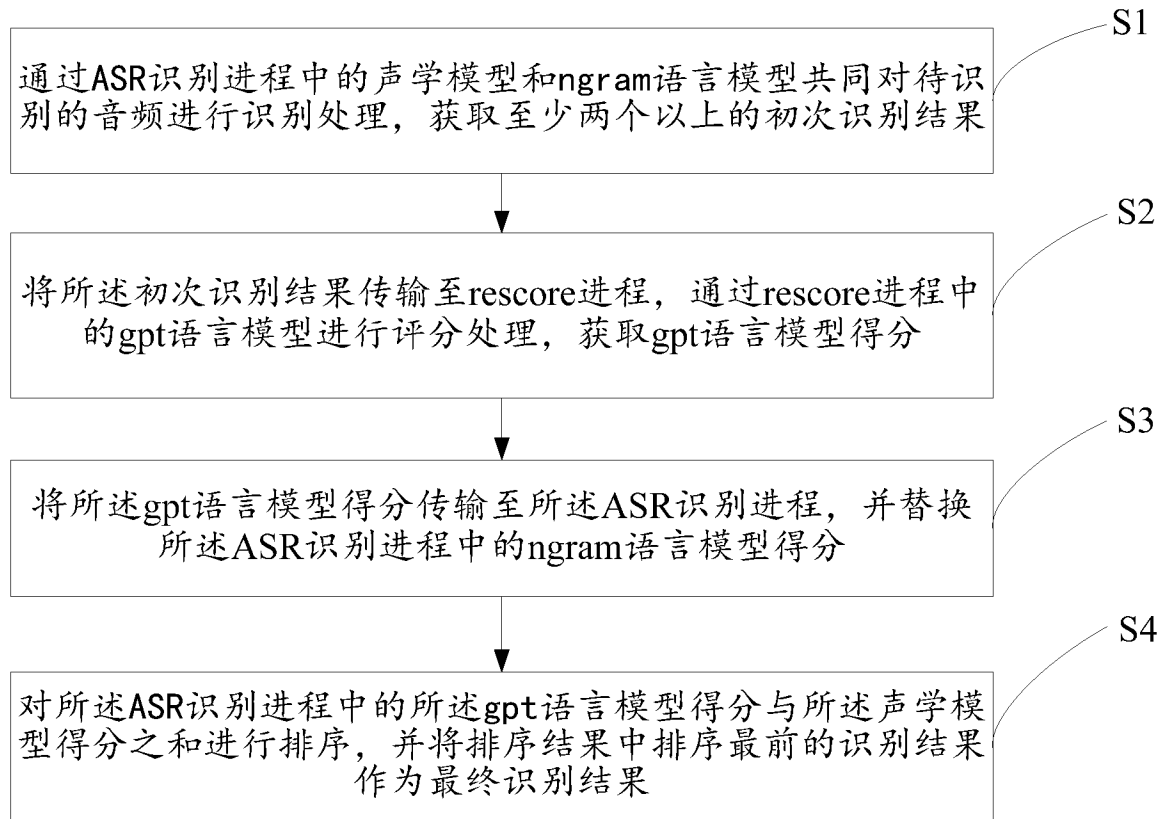


图 1

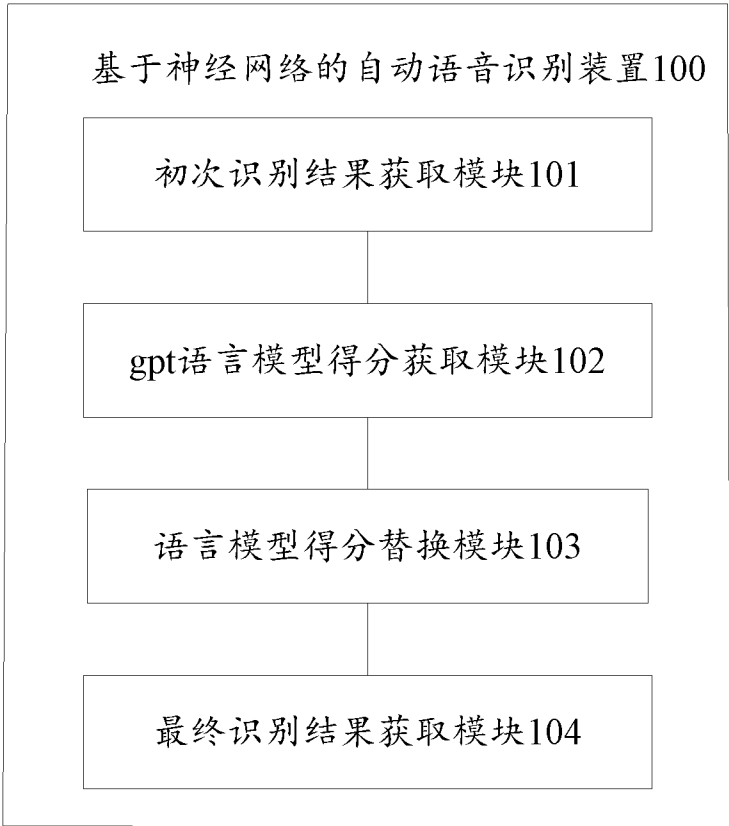


图 2

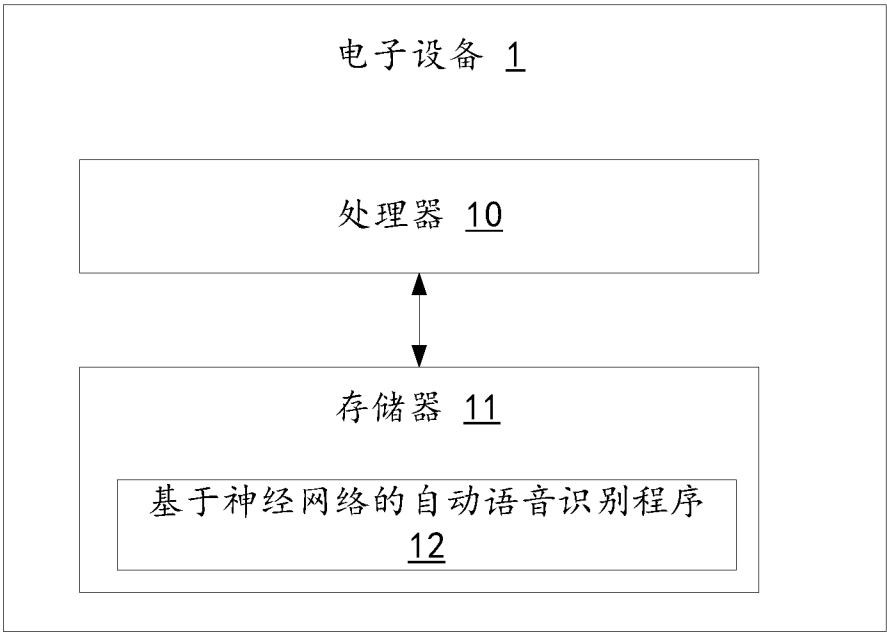


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/071220

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/32(2013.01)i; G10L 15/06(2013.01)i; G10L 15/14(2006.01)i; G10L 15/16(2006.01)i; G06N 3/02(2006.01)i; G06K 9/62(2022.01)i; G06F 40/284(2020.01)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; G06N; G06K; 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)

CNABS; CNTXT; CNKI; VEN; USTXT; WOTXT; EPTXT; IEEE: 平安科技, 方明, 魏韬, 马骏, 王少军, 声学模型, 语言模型, n元, 得分, 替换, 代替, 替代, 更换, 重打分, 分, 值, 替, 代, 换, 之和, 合, 累积, 累计, 汇总, 总分, 总值, 音频, 帧, 后验概率, 时间边界, acoustic model, language model, ngram, n-gram, gram, gpt, rescore, score?, replaced, instead of, neural network, sum, order+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113450805 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 28 September 2021 (2021-09-28) claims 1-10, and description, paragraphs [0002]-[0154], and figures 1-3	1-20
X	CN 112562640 A (BEIJING SOUNDAL TECHNOLOGY CO., LTD.) 26 March 2021 (2021-03-26) description, paragraphs [0040]-[0094], and figures 1-5	1-3, 5-11, 13-17, 19, 20
Y	CN 112562640 A (BEIJING SOUNDAL TECHNOLOGY CO., LTD.) 26 March 2021 (2021-03-26) description, paragraphs [0040]-[0094], and figures 1-5	4, 12, 18
Y	CN 104575490 A (SUZHOU CHIVOX INFORMATION TECHNOLOGY CO., LTD.) 29 April 2015 (2015-04-29) description, paragraphs [0041]-[0054], and figure 1	4, 12, 18
A	CN 110797026 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 14 February 2020 (2020-02-14) 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

02 March 2022

Date of mailing of the international search report

17 March 2022

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

International application No.

PCT/CN2022/071220**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112349289 A (BEIJING SINOVOICE TECHNOLOGY CO., LTD.) 09 February 2021 (2021-02-09) entire document	1-20
A	US 2015179169 A1 (JOHN, V. G. et al.) 25 June 2015 (2015-06-25) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/071220

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113450805	A	28 September 2021	None			
CN	112562640	A	26 March 2021	None			
CN	104575490	A	29 April 2015	CN	104575490	B	07 November 2017
CN	110797026	A	14 February 2020	HK	40021011	A0	30 October 2020
CN	112349289	A	09 February 2021	None			
US	2015179169	A1	25 June 2015	None			

国际检索报告

国际申请号

PCT/CN2022/071220

A. 主题的分类 G10L 15/32(2013.01)i; G10L 15/06(2013.01)i; G10L 15/14(2006.01)i; G10L 15/16(2006.01)i; G06N 3/02(2006.01)i; G06K 9/62(2022.01)i; G06F 40/284(2020.01)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; G06N; G06K; 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)) CNABS; CNTXT; CNKI; VEN; USTXT; WOTXT; EPTXT; IEEE: 平安科技, 方明, 魏韬, 马骏, 王少军, 声学模型, 语言模型, n元, 得分, 替换, 代替, 替代, 更换, 重打分, 分, 值, 替, 代, 换, 之和, 合, 累积, 累计, 汇总, 总分, 总值, 音频, 帧, 后验概率, 时间边界, acoustic model, language model, ngram, n-gram, gram, gpt, rescore, score?, replaced, instead of, neural network, sum, order+				
C. 相关文件				
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求		
PX	CN 113450805 A (平安科技深圳有限公司) 2021年9月28日 (2021 - 09 - 28) 权利要求1-10, 说明书第[0002]-[0154]段, 附图1-3	1-20		
X	CN 112562640 A (北京声智科技有限公司) 2021年3月26日 (2021 - 03 - 26) 说明书第[0040]-[0094]段, 附图1-5	1-3、5-11、 13-17、19、20		
Y	CN 112562640 A (北京声智科技有限公司) 2021年3月26日 (2021 - 03 - 26) 说明书第[0040]-[0094]段, 附图1-5	4、12、18		
Y	CN 104575490 A (苏州驰声信息科技有限公司) 2015年4月29日 (2015 - 04 - 29) 说明书第[0041]-[0054]段, 附图1	4、12、18		
A	CN 110797026 A (腾讯科技深圳有限公司) 2020年2月14日 (2020 - 02 - 14) 全文	1-20		
A	CN 112349289 A (北京捷通华声科技股份有限公司) 2021年2月9日 (2021 - 02 - 09) 全文	1-20		
A	US 2015179169 A1 (JOHN VIJAY GEORGE 等) 2015年6月25日 (2015 - 06 - 25) 全文	1-20		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。				
<table><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 2022年3月2日		国际检索报告邮寄日期 2022年3月17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 张鲁鲁 电话号码 (86-512) 88997385		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/071220

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	113450805	A	2021年9月28日	无	
CN	112562640	A	2021年3月26日	无	
CN	104575490	A	2015年4月29日	CN 104575490 B	2017年11月7日
CN	110797026	A	2020年2月14日	HK 40021011 A0	2020年10月30日
CN	112349289	A	2021年2月9日	无	
US	2015179169	A1	2015年6月25日	无	