

(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 2020/098100 A1

(43) 国际公布日
2020 年 5 月 22 日 (22.05.2020)

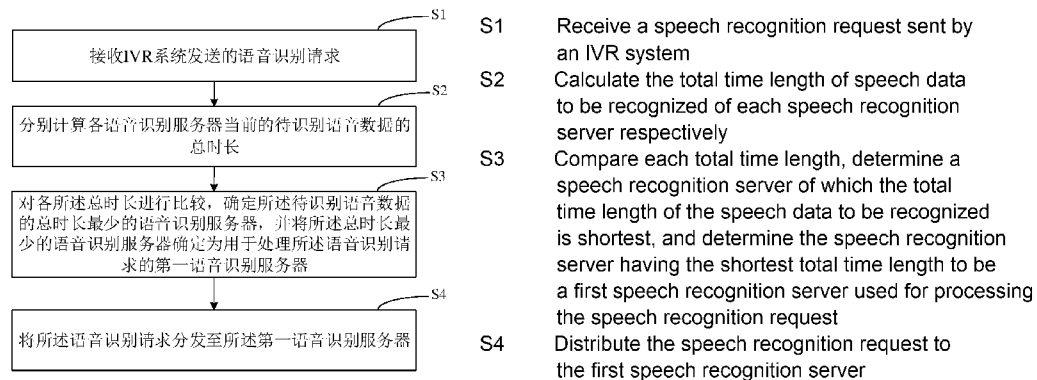
- (51) 国际专利分类号:
H04L 29/08 (2006.01) *G10L 15/30* (2013.01)
- (21) 国际申请号: PCT/CN2018/124400
- (22) 国际申请日: 2018 年 12 月 27 日 (27.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811340885.4 2018 年 11 月 12 日 (12.11.2018) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY(SHENZHEN)CO.,LTD.) [CN/CN];
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: 张垒(ZHANG, Lei); 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)。 邹芳(ZOU, Fang); 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)。 邢艳(XING, Yan); 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)。 李晋(LI, Jin); 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)。 张裕生(ZHANG, Yusheng); 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事务所(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智恒新兴产业园 E 区 01B 栋 405 室, Guangdong 518000 (CN)。

(54) Title: METHOD AND APPARATUS FOR ALLOCATING RESOURCES ON BASIS OF DATA ANALYSIS, AND COMPUTER DEVICE

(54) 发明名称: 基于数据分析的资源分配方法、装置和计算机设备



(57) Abstract: Provided in the present application are a method and apparatus for allocating resources on the basis of data analysis, and a computer device: receiving a speech recognition request sent by an IVR system; calculating the total time length of speech data to be recognized of each speech recognition server; comparing the total time lengths, and determining the speech recognition server having the shortest total time length to be a first speech recognition server for processing the speech recognition request; and distributing the speech recognition request to the first speech recognition server. The present application improves the processing efficiency for speech recognition requests.

(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, 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))。

(57) 摘要: 本申请揭示了基于数据分析的资源分配方法、装置和计算机设备, 接收IVR系统发送的语音识别请求; 计算各语音识别服务器的待识别语音数据的总时长; 对总时长进行比较, 将总时长最少的语音识别服务器确定为处理语音识别请求的第一语音识别服务器; 将语音识别请求分发至第一语音识别服务器。本申请提高了语音识别请求的处理效率。

基于数据分析的资源分配方法、装置和计算机设备

[0001] 本申请要求于2018年11月12日提交中国专利局、申请号为201811340885.4，发明名称为“基于数据分析的资源分配方法、装置和计算机设备”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] 本申请涉及语音识别领域，具体涉及一种基于数据分析的资源分配方法、装置和计算机设备。

背景技术

[0003] 语音识别技术，是指让机器通过识别和理解，将语音信号转换为相应的文本或命令的技术，也就是说，要让机器听懂人类的语音。

[0004] 负载均衡器用于将接收到的语音识别请求分发至语音识别服务器，语音识别服务器用于对接收到的语音识别请求进行处理。现有技术中，在接收到IVR(Interactive Voice Response，交互式语音应答)系统所发出的语音识别请求后，负载均衡器会使用轮询的方式来进行对语音识别服务器的调用，即选择多个语音识别服务器中的一个来处理接收到的语音识别请求，以实现各个语音识别服务器的负载均衡。但轮询只能根据交互次数来进行分配，且每次交互的时长可能存在较大的差异，由于负载均衡器没有根据各语音识别服务器的资源的实际使用情况来选择合理的语音识别服务器处理进入的语音识别请求，容易导致各语音识别服务器的资源分配不平均，语音识别请求的处理效率不高。

[0005] 因此，现有技术还存在缺陷，需要改进。

发明概述

技术问题

[0006] 本申请的主要目的为提供一种基于数据分析的资源分配方法、装置和计算机设备，旨在解决现有对于语音识别服务器的资源分配不平均，语音识别请求的处理效率不高的技术问题。

问题的解决方案

技术方案

- [0007] 为了实现上述申请目的，本申请提出一种基于数据分析的资源分配方法，包括：
- [0008] 接收IVR（交互式语音应答）系统发送的语音识别请求；
- [0009] 分别计算各语音识别服务器当前的待识别语音数据的总时长；
- [0010] 对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
- [0011] 将所述语音识别请求分发至所述第一语音识别服务器。
- [0012] 本申请还提供了一种基于数据分析的资源分配装置，包括：
- [0013] 第一接收模块，用于接收IVR系统发送的语音识别请求；
- [0014] 第一计算模块，用于分别计算各语音识别服务器当前的待识别语音数据的总时长；
- [0015] 第一比较模块，用于对各所述总时长进行比较，得出所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
- [0016] 第一分发模块，用于将所述识别请求分发至所述第一语音识别服务器。
- [0017] 本申请还提供一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现上述方法的步骤。
- [0018] 本申请还提供一种计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述方法的步骤。

发明的有益效果

有益效果

- [0019] 本申请中提供的基于数据分析的资源分配方法、装置、计算机设备和存储介质，接收IVR系统发送的语音识别请求；分别计算各语音识别服务器当前的待识别语音数据的总时长；对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；将所述识别请求分发至所述

第一语音识别服务器。本申请通过对各语音识别服务器当前的待识别语音数据的总时长进行比较分析，来选择合理的语音识别服务器处理接收到的语音识别请求，有利于实现对语音识别服务器的资源均衡使用，提高语音识别请求的处理效率。

对附图的简要说明

附图说明

- [0020] 图1是本申请一实施例的基于数据分析的资源分配方法的流程示意图；
- [0021] 图2是本申请另一实施例的基于数据分析的资源分配方法的流程示意图；
- [0022] 图3是本申请一实施例的基于数据分析的资源分配装置的结构示意框图；
- [0023] 图4是本申请一实施例的计算机设备的结构示意框图。

实施该发明的最佳实施例

本发明的最佳实施方式

- [0024] 应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。
- [0025] 参照图1，本申请一实施例中提供了一种基于数据分析的资源分配方法，包括以下步骤：
 - [0026] S1：接收IVR（交互式语音应答）系统发送的语音识别请求；
 - [0027] S2：分别计算各语音识别服务器当前的待识别语音数据的总时长；
 - [0028] S3：对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
 - [0029] S4：将所述语音识别请求分发至所述第一语音识别服务器。
- [0030] 如上述步骤S1所述，本实施例的执行主体为基于数据分析的资源分配装置，具体可为负载均衡器。负载均衡器用于接收IVR(Interactive Voice Response，交互式语音应答)系统发送的语音识别请求，并将该语音识别请求分发至语音识别服务器，以使得语音识别服务器对该语音识别请求进行处理。其中，在接收IVR系统发送的语音识别请求之前，负载均衡器需要与IVR系统建立对接，以实现与IVR系统的信息交互。举例地，负载均衡器与IVR系统建立对接的方式可为：在负

载均衡器的sdp消息中带上识别资源信息，来建立负载均衡器与IVR系统的对接，其中sdp消息为IVR系统建立会话连接时使用的消息体；或者在IVR系统内直接配置负载池，使得IVR系统能通过该负载池直接与负载均衡器建立对接。

[0031] 如上述步骤S2至S4所述，在接收到IVR系统发送的语音识别请求后，负载均衡器会对各语音识别服务器的资源实际使用率进行分析。其中，上述资源实际使用率通过各语音识别服务器的待识别语音数据的时长数据，即通过分析计算出的各所述语音识别服务器当前的待识别语音数据的总时长来体现，由于负载均衡器记录有由IVR系统发送的所有的语音识别请求的数据信息，该数据信息包括各个语音数据的时长，以及将各个语音数据对应发送的各个语音识别服务器，通过将每个语音识别服务器所需要处理的语音数据的时长数据减去已经返回到负载均衡器的已处理语音数据时长数据，得到的差值就是每个语音识别服务器当前的待处理的语音数据的总时长，通过对各语音识别服务器当前的待识别语音数据的总时长进行比较分析，可以确定出待识别语音数据的总时长最少的语音识别服务器，并将该总时长最少的语音识别服务器，确定为用于处理该语音识别请求的第一语音识别服务器。现有技术中，语音识别服务器的资源使用率是通过语音数据的数量来表现的，举例地，语音识别服务器A1有2条待识别语音数据，语音识别服务器A2有5条待识别语音数据，当收到语音识别请求时，负载均衡器会将该语音识别请求分发到待识别语音数据的数量较少的语音识别服务器A1，但负载均衡器并不会分析语音识别服务器A1和A2中待识别语音数据的时长，假如语音识别服务器A1的2条待识别语音数据的总时长是5分钟，语音识别服务器A2的5条待识别语音数据的总时长是2分钟，那么语音识别服务器A1的资源实际使用率是大于语音识别服务器A2的，可见现有的对语音识别服务器的资源分配的方法并不合理，此时应该将该识别请求分发到资源实际使用率较小的语音识别服务器A2，才更有利于实现语音识别服务器的资源均衡分配。本实施例利用待识别语音数据的总时长来表征资源实际使用率，能准确地体现出当前各语音识别服务器的资源的实际使用情况，并根据当前各语音识别服务器的资源的实际使用情况，来将语音识别请求合理的分发到资源实际使用率较低的语音识别服务器，有利于实现对语音识别服务器的资源均衡使用，提高语音识别

请求的处理效率。

[0032] 参照图2, 进一步地, 本申请一实施例中, 上述步骤S4之后, 包括:

[0033] S400: 判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器;

[0034] S401: 若是, 将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。

[0035] 本实施例中, 预先设置的预设时长阈值, 代表着语音识别服务器所能支持处理的最大的语音数据的时长数, 如果该语音识别服务器在长时间内处理的语音数据的时长数都大于该预设时长阈值, 即意味着该语音识别服务器长时间处于超负荷工作状态, 则该语音识别服务器会受到严重的损坏甚至报废。为了保证语音识别服务器能够保持良好的使用寿命, 负载均衡器需要时刻监控各个语音识别服务器的工作状态, 当发现存在待识别语音数据的第一总时长大于预设时长阈值的第二语音识别服务器, 则会将所述第二语音识别服务器移入暂不接收新的语音识别请求的黑名单内, 即在一段时间内不会向该第二语音识别服务器分发新的语音识别请求, 只有当第二语音识别服务器内的待处理的语音数据的总时长小于预设时长阈值后, 才会将第二语音识别服务器从该黑名单内移除。在本实施例, 当负载均衡器监控到某语音识别服务器处于超负荷工作状态时, 会智能地停止向该语音识别服务器分发新的语音识别请求, 来有效地对该语音识别服务器进行保护, 保证了该语音识别服务器的使用寿命。

[0036] 进一步地, 本申请一实施例中, 上述步骤S401之后, 包括:

[0037] S402: 根据各所述语音识别服务器的待识别语音数据的总时长, 计算出各所述语音识别服务器的待识别语音数据的平均时长;

[0038] S403: 筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器;

[0039] S404: 经过第一预设时间后, 判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值;

[0040] S405: 若是, 将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器, 以使得经过分配后的所述第二语音识别服务器内的待识

别语音数据的总时长小于所述预设时长阈值。

[0041] 如上述步骤S402至S403所述，根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长的步骤，具体为：对当前所有的语音识别服务器的待识别语音数据的总时长进行求和运算得到和值，再将该和值除以所有的语音识别服务器的数量，便能得到该平均时长，进一步地，可筛选出待识别语音数据的总时长小于所述平均时长的语音识别服务器，其中，语音识别服务器的数量由实际情况决定。

[0042] 如上述步骤S404至S405所述，预先设置的第一预设时间，表示语音识别处理器在该第一预设时间内处于超负荷工作状态，几乎不会对语音识别处理器本身造成损坏。在本实施例，在第二语音识别服务器对待识别语音数据进行处理，如果在经过第一预设时间后，第二语音识别服务器内剩余的待识别语音数据的总时长小于上述预设时长阈值，则表明第二语音识别服务器并没有长时间处于超负荷工作状态；而如果在超过该第一预设时间后，第二语音识别服务器内剩余的待识别语音数据的总时长仍然大于上述预设时长阈值，则表明第二语音识别服务器长时间处于超负荷工作状态，这样会对第二语音识别服务器本身的使用寿命造成极大的影响。为了对第二语音识别服务器进行保护，保证语音识别服务器的使用寿命，此时需要将第二语音识别处理器内的部分未识别语音数据分配至其它语音识别服务器，例如第三语音识别服务器进行处理，该第三语音识别服务器为待识别语音数据的总时长小于该平均时长的语音识别服务器，其中，上述部分待识别语音数据可为第二语音识别服务器中所有的待识别语音数据中的任意部分的待识别语音数据，需要进行分配的部分待识别语音数据的时长大于或等于指定时长，上述指定时长为经过第一预设时间后的所述第二语音识别服务器内的待识别语音数据的总时长与上述预设时长阈值的差值。另外，将第二语音识别服务器内的部分待识别语音数据分配至待识别语音数据的总时长小于平均时长的第三语音识别服务器，第三语音识别服务器的数量根据实际情况而定，具体为：对第二语音识别服务器内的语音数据的实际时长数据进行分析，假如该实际时长数据不会很大，则可将语音识别服务器内的部分待识别语音数据只分配至待识别语音数据的总时长小于平均时长的某一个指定语音识别

服务器，并且保证第二语音识别服务器在经过分配后，剩下的待识别语音数据能在经过该第一预设时间的处理后的总时长小于该预设时长阈值；或者假如该实际时长数据过大，则可将第二语音识别服务器内的部分待识别语音数据分配到待识别语音数据的总时长小于所述平均时长的多个指定语音识别服务器（具体数量根据实际情况设置），并且保证第二语音识别服务器在经过分配后，剩下的待识别语音数据能在经过该第一预设时间的处理后的总时长能小于该预设时长阈值。通过将负荷量大于预设值的语音识别服务器内的部分待识别语音数据分配到负荷量较小的一个或多个指定语音识别服务器，有效地保证了负荷量大的语音识别服务器的使用寿命，也进一步促进了各语音识别服务器的资源的均衡使用。

[0043] 进一步地，本申请一实施例中，上述步骤S4之后，包括：

[0044] S410：判断所述第一语音识别服务器是否处于正常工作状态；

[0045] S411：若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；

[0046] S412：根据所述比较结果，确定出用于处理所述识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；

[0047] S413：将所述语音识别请求分发至所述第四语音识别服务器。

[0048] 本实施例中，在负载均衡器确定了将所述语音识别请求分发至第一语音识别服务器后，可能会出现第一语音识别服务器处于无法正常工作状态的情况，那么第一语音识别服务器就会无法对该语音识别请求进行响应反馈，以及后续的对与该语音识别请求相对应的语音数据的识别工作，使得整个识别流程无法正常运转，IVR系统不能收到第一语音识别服务器的识别结果，且用户无法及时得到IVR系统的回应，造成用户使用体验差。所以在本实施例，需要进一步确定第一语音识别服务器是否正常工作状态，如果判定第一语音识别服务器处于无法正常工作的状态，则将第一语音识别服务器排除掉，并对所述第一语音识别服务器之外的其它语音识别服务器的资源实际使用率进行分析，确定出用于处理所述语音识别请求的第四语音识别服务器，其中第四语音识别服务器当前的待识别语音数据的总时长在上述其它语音识别服务器之中是最少的，并在确认第四

语音识别服务器处于正常工作状态后，将所述语音识别请求分发至所述第四语音识别服务器，以便该第四语音识别服务器进行对应处理。在本实施例，当第一语音识别服务器无法正常工作时，通过调用处于正常工作状态的第四语音识别服务器来对语音识别请求进行处理，有利于识别流程的正常运作，IVR系统能及时收到第四语音识别服务器的识别结果，且用户也能及时得到IVR系统的回应，保证了用户的使用体验。

[0049] 进一步地，本申请一实施例中，上述步骤S412，包括：

[0050] S4120：按照待识别语音数据的总时长从小到大的顺序对所述其它语音识别服务器进行排序；

[0051] S4121：根据所述排序，遍历所述其它语音识别服务器；

[0052] S4122：判断当前访问的语音识别服务器是否处于正常工作状态；

[0053] S4123：若是，将所述当前访问的语音识别服务器确定为用于处理所述语音识别请求的所述第四语音识别服务器，并停止遍历。

[0054] 本实施例中，根据上述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器的步骤，具体包括：当第一语音识别服务器为处于无法正常工作状态，则通过对除去第一语音识别服务器之外的其它各语音识别服务器的总时长进行分析，按照待识别语音数据的总时长从小到大的顺序，即递增的方式对其它各语音识别服务器进行排序，并根据该排序遍历上述的其它语音识别服务器，在每遍历到一个语音识别服务器时，如果确定该语音识别服务器处于正常工作状态，则将该语音识别服务器确定为用于处理所述语音识别请求的上述第四语音识别服务器，并将该语音识别请求分发至该语音识别服务器进行处理，并停止遍历。进一步地，如果在遍历结束后，都没有找到处于正常工作状态的语音识别服务器，则会向IVR系统返回处理失败的信息，即当前没有正常可用的语音识别服务器对上述语音识别请求进行处理，另外还会发出维修信息，使得维修人员对处于无法正常工作状态的语音识别服务器进行维修。

[0055] 进一步地，本申请一实施例中，上述步骤S1之前，包括：

[0056] S100：与各所述语音识别服务器建立TCP（传输控制协议）长连接，以便通过所述TCP长连接与各所述语音识别服务器之间进行信息交互。

[0057] 本实施例中，在负载均衡器进行初始化时，可与每个语音识别服务器建立的TCP（Transmission Control Protocol，传输控制协议）长连接，可选地，可与每一个语音识别服务器建立数量至少为1的TCP长连接，其中建立的TCP长连接的具体数量按实际情况进行确定。在该传输控制协议长连接成功建立后，当负载均衡器需要将IVR系统发送的关于某一语音数据的语音识别请求分发至某一语音识别服务器时，即可直接使用上述所建立的传输控制协议长连接，通过所述传输控制协议长连接与该语音识别服务器之间进行信息交互，从而省去了当需要负载均衡器来调用该语音识别服务器时，再去建立传输控制协议长连接所耗费的时间。另外，负载均衡器与每个语音识别服务器之间所建立的传输控制协议长连接的个数的具体取值可根据实际需要而定，可以为一个，也可以为多个，举例地，负载均衡器与语音识别服务器A建立了一个传输控制协议长连接，负载均衡器与语音识别服务器B建立了两个传输控制协议长连接，负载均衡器与语音识别服务器C建立了五个传输控制协议长连接。其中，建立多个传输控制协议长连接的好处在于：当负载均衡器同时接收到IVR系统发送的多个语音数据对应的多个语音识别请求，且确定多个语音识别请求均需要由同一个语音识别服务器进行处理时，可利用多个传输控制协议长连接分别将多个语音识别请求同时分发至该语音识别服务器，而如果只有一个传输控制协议长连接的话，则只能在分发完一个语音识别请求后再分发另一个语音识别请求，直至将该多个语音识别请求分发完毕，所以建立一定数量的传输控制协议长连接，能够有效的提高了语音识别请求的传输效率。

[0058] 进一步地，本申请一实施例中，上述步骤S4之后，包括：

[0059] S420：判断是否接收到所述第一语音识别服务器反馈的对于所述语音识别请求的响应信息；

[0060] S421：若是，将所述响应信息转发至所述IVR系统；

[0061] S422：接收由所述IVR系统发送的与所述语音识别请求对应的语音数据；

[0062] S423：将所述语音数据发送至所述第一语音识别服务器，以便所述第一语音识别服务器对所述语音数据进行识别。

[0063] 本实施例中，在负载均衡器将所述语音识别请求分发到第一语音识别服务器后

，第一语音识别服务器会对该语音识别请求的语法进行解析，如果解析成功则会向负载均衡器返回请求识别成功的响应信息，而如果解析失败，则会向负载均衡器返回错误。在负载均衡器接收到上述响应信息后，会将该响应信息转发至上述的IVR系统，IVR系统在接收到响应信息后会向负载均衡器发送与上述语音识别请求相对应的语音数据，在负载均衡器接收到该语音数据后便会将其转发到该第一语音识别服务器，之后第一语音识别服务器会根据语法，使用对应的语言模型对该语音数据进行识别，得到识别结果，并将该识别结果返回到负载均衡器，使得负载均衡器将该识别结果返回到IVR系统，以实现IVR系统根据该识别结果与用户进行信息交互。

[0064] 进一步地，本申请一实施例中，在接收IVR系统发送的语音识别请求的步骤之前，IVR系统会向负载均衡器发出心跳信息，负载均衡器在接收到所述IVR系统发送的心跳信息后，会判断自身是否能够发送对于所述心跳信息的反馈信息，若否，负载均衡器会发出通过第二负载均衡器接收IVR发送的语音数据的识别请求的警报信息。本实施例中，上述负载均衡器为主用负载均衡器，并对应设置至少一个备用负载均衡器，即上述第二负载均衡器。当上述负载均衡器，即主用负载均衡器在接收所述IVR系统发送的心跳信息，并判定自身出现故障，无法发送对于所述心跳信息的反馈信息后，会自动发出警报信息，以请求快速获得对应的维修，另外，主用负载均衡器也会发出通过备用负载均衡器接收IVR系统发送的语音识别请求的警报信息，在IVR系统得知主用负载均衡器出现故障，IVR会将语音识别请求发送至备用负载均衡器，备用负载均衡器会代替主用负载均衡器工作，来接收IVR发送的语音数据的识别请求，并进行后续的处理，并对当前各语音识别服务器的资源实际使用率分析，确定出用于处理所述识别请求的第一语音识别服务器，并将所述识别请求分发至所述第一语音识别服务器。通过增加备用负载均衡器，有效地提高了负载均衡器的可用性，保证服务的连续、稳定运行。另外，备用负载均衡器也可主动检测主用负载均衡器是否发送出了对于心跳信息的反馈信息，当判定主用负载均衡器出了故障，即无法正常工作后，可直接代替主用负载均衡器工作，来提醒IVR系统将语音数据的识别请求发送至备用负载均衡器。

- [0065] 本申请实施例的基于数据分析的资源分配方法，通过对各语音识别服务器当前的待识别语音数据的总时长进行比较分析，来选择合理的语音识别服务器处理接收到的语音识别请求，有利于实现对语音识别服务器的资源均衡使用，提高语音识别请求的处理效率。
- [0066] 参照图3，本申请一实施例中还提供了一种基于数据分析的资源分配装置，包括：
- [0067] 第一接收模块1，用于接收IVR（交互式语音应答）系统发送的语音识别请求；
- [0068] 第一计算模块2，分别计算各语音识别服务器当前的待识别语音数据的总时长；
- [0069] 第一比较模块3，用于对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
- [0070] 第一分发模块4，用于将所述语音识别请求分发至所述第一语音识别服务器。
- [0071] 在上述第一接收模块中，本实施例的执行主体为基于数据分析的资源分配装置，具体可为负载均衡器。负载均衡器用于接收IVR(Interactive Voice Response，交互式语音应答)系统发送的语音识别请求，并将该语音识别请求分发至语音识别服务器，以使得语音识别服务器对该语音识别请求进行处理。其中，在接收IVR系统发送的语音识别请求之前，负载均衡器需要与IVR系统建立对接，以实现与IVR系统的信息交互。举例地，负载均衡器与IVR系统建立对接的方式可为：在负载均衡器的sdp消息中带上识别资源信息，来建立负载均衡器与IVR系统的对接，其中sdp消息为IVR系统建立会话连接时使用的消息体；或者在IVR系统内直接配置负载池，使得IVR系统能通过该负载池直接与负载均衡器建立对接。
- [0072] 在上述第一计算模块、第一比较模块和第一分发模块中，在接收到IVR系统发送的语音识别请求后，负载均衡器会对各语音识别服务器的资源实际使用率进行分析。其中，上述资源实际使用率通过各语音识别服务器的待识别语音数据的时长数据，即通过分析计算出的各所述语音识别服务器当前的待识别语音数据的总时长来体现，由于负载均衡器记录有由IVR系统发送的所有的语音识别请求的数据信息，该数据信息包括各个语音数据的时长，以及将各个语音数据对

应发送的各个语音识别服务器，通过将每个语音识别服务器所需要处理的语音数据的时长数据减去已经返回到负载均衡器的已处理语音数据时长数据，得到的差值就是每个语音识别服务器当前的待处理的语音数据的总时长，通过对各语音识别服务器当前的待识别语音数据的总时长进行比较分析，可以确定出待识别语音数据的总时长最少的语音识别服务器，并将该总时长最少的语音识别服务器，确定为用于处理该语音识别请求的第一语音识别服务器。现有技术中，语音识别服务器的资源使用率是通过语音数据的数量来表现的，举例地，语音识别服务器A1有2条待识别语音数据，语音识别服务器A2有5条待识别语音数据，当收到语音识别请求时，负载均衡器会将该语音识别请求分发到待识别语音数据的数量较少的语音识别服务器A1，但负载均衡器并不会分析语音识别服务器A1和A2中待识别语音数据的时长，假如语音识别服务器A1的2条待识别语音数据的总时长是5分钟，语音识别服务器A2的5条待识别语音数据的总时长是2分钟，那么语音识别服务器A1的资源实际使用率是大于语音识别服务器A2的，可见现有的对语音识别服务器的资源分配的方法并不合理，此时应该将该识别请求分发到资源实际使用率较小的语音识别服务器A2，才更有利于实现语音识别服务器的资源均衡分配。本实施例利用待识别语音数据的总时长来表征资源实际使用率，能准确地体现出当前各语音识别服务器的资源的实际使用情况，并根据当前各语音识别服务器的资源的实际使用情况，来将语音识别请求合理的分发到资源实际使用率较低的语音识别服务器，有利于实现对语音识别服务器的资源均衡使用，提高语音识别请求的处理效率。

[0073] 进一步地，本申请一实施例中，上述基于数据分析的资源分配装置，包括：

[0074] 第一判断模块，用于判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器；

[0075] 加入模块，用于若是，将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。

[0076] 本实施例中，预先设置的预设时长阈值，代表着语音识别服务器所能支持处理的最大的语音数据的时长数，如果该语音识别服务器在长时间内处理的语音数据的时长数都大于该预设时长阈值，即意味着该语音识别服务器长时间处于超

负荷工作状态，则该语音识别服务器会受到严重的损坏甚至报废。为了保证语音识别服务器能够保持良好的使用寿命，负载均衡器需要时刻监控各个语音识别服务器的工作状态，当发现存在待识别语音数据的第一总时长大于预设时长阈值的第二语音识别服务器，则会将所述第二语音识别服务器移入暂不接收新的语音识别请求的黑名单内，即在一段时间内不会向该第二语音识别服务器分发新的语音识别请求，只有当第二语音识别服务器内的待处理的语音数据的总时长小于预设时长阈值后，才会将第二语音识别服务器从该黑名单内移除。在本实施例，当负载均衡器监控到某语音识别服务器处于超负荷工作状态时，会智能地停止向该语音识别服务器分发新的语音识别请求，来有效地对该语音识别服务器进行保护，保证了该语音识别服务器的使用寿命。

[0077] 进一步地，本申请一实施例中，上述基于数据分析的资源分配装置，包括：

[0078] 第二计算模块，用于根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长；

[0079] 筛选模块，用于筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器；

[0080] 第二判断模块，用于经过第一预设时间后，判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值；

[0081] 分配模块，用于若是，将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器，以使得经过分配后的所述第二语音识别服务器内的待识别语音数据的总时长小于所述预设时长阈值。

[0082] 在上述第二计算模块和筛选模块中，根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长的步骤，具体为：对当前所有的语音识别服务器的待识别语音数据的总时长进行求和运算得到和值，再将该和值除以所有的语音识别服务器的数量，便能得到该平均时长，进一步地，可筛选出待识别语音数据的总时长小于所述平均时长的语音识别服务器，其中，语音识别服务器的数量由实际情况决定。

[0083] 在上述第二判断模块和分配模块中，预先设置的第一预设时间，表示语音识别处理器在该第一预设时间内处于超负荷工作状态，几乎不会对语音识别处理器

本身造成损坏。在本实施例，在第二语音识别服务器对待识别语音数据进行处理，如果在经过第一预设时间后，第二语音识别服务器内剩余的待识别语音数据的总时长小于上述预设时长阈值，则表明第二语音识别服务器并没有长时间处于超负荷工作状态；而如果在超过该第一预设时间后，第二语音识别服务器内剩余的待识别语音数据的总时长仍然大于上述预设时长阈值，则表明第二语音识别服务器长时间处于超负荷工作状态，这样会对第二语音识别服务器本身的使用寿命造成极大的影响。为了对第二语音识别服务器进行保护，保证语音识别服务器的使用寿命，此时需要将第二语音识别处理器内的部分未识别语音数据分配到其他语音识别服务器，例如第三语音识别服务器进行处理，该第三语音识别服务器为待识别语音数据的总时长小于该平均时长的语音识别服务器，其中，上述部分待识别语音数据可为第二语音识别服务器中所有的待识别语音数据中的任意部分的待识别语音数据，需要进行分配的部分待识别语音数据的时长大于或等于指定时长，上述指定时长为经过第一预设时间后的所述第二语音识别服务器内的待识别语音数据的总时长与所述预设时长阈值的差值。另外，将第二语音识别服务器内的部分待识别语音数据分配至待识别语音数据的总时长小于平均时长的第三语音识别服务器，第三语音识别服务器的数量根据实际情况而定，具体为：对第二语音识别服务器内的语音数据的实际时长数据进行分析，假如该实际时长数据不会很大，则可将语音识别服务器内的部分待识别语音数据只分配到待识别语音数据的总时长小于平均时长的某一个指定语音识别服务器，并且保证第二语音识别服务器在经过分配后，剩下的待识别语音数据能在经过该第一预设时间的处理后的总时长小于该预设时长阈值；或者假如该实际时长数据过大，则可将第二语音识别服务器内的部分待识别语音数据分配到待识别语音数据的总时长小于所述平均时长的多个指定语音识别服务器（具体数量根据实际情况设置），并且保证第二语音识别服务器在经过分配后，剩下的待识别语音数据能在经过该第一预设时间的处理后的总时长能小于该预设时长阈值。通过将负荷量大于预设值的语音识别服务器内的部分待识别语音数据分配到负荷量较小的一个或多个指定语音识别服务器，有效地保证了负荷量大的语音识别服务器的使用寿命，也进一步促进了各语音识别服务器的

资源的均衡使用。

[0084] 进一步地，本申请一实施例中，上述基于数据分析的资源分配装置，包括：

[0085] 第三判断模块，用于判断所述第一语音识别服务器是否处于正常工作状态；

[0086] 第二比较模块，用于若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；

[0087] 确定模块，用于根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；

[0088] 第二分发模块，用于将所述语音识别请求分发至所述第四语音识别服务器。

[0089] 本实施例中，在负载均衡器确定了将所述语音识别请求分发至第一语音识别服务器后，可能会出现第一语音识别服务器处于无法正常工作状态的情况，那么第一语音识别服务器就会无法对该语音识别请求进行响应反馈，以及后续的对与该语音识别请求相对应的语音数据的识别工作，使得整个流程无法正常运作，IVR系统不能收到第一语音识别服务器的识别结果，且用户无法及时得到IVR系统的回应，造成用户使用体验差。所以在本实施例，需要进一步确定第一语音识别服务器是否正常工作状态，如果判定第一语音识别服务器处于无法正常工作的状态，则将第一语音识别服务器排除掉，并对所述第一语音识别服务器之外的其它语音识别服务器的资源实际使用率进行分析，确定出用于处理所述语音识别请求的第四语音识别服务器，其中第四语音识别服务器当前的待识别语音数据的总时长在上述其它语音识别服务器之中是最少的，并在确认第四语音识别服务器处于正常工作状态后，将所述语音识别请求分发至所述第四语音识别服务器，以便该第四语音识别服务器进行对应处理。在本实施例，当第一语音识别服务器无法正常工作时，通过调用处于正常工作状态的第四语音识别服务器来对语音识别请求进行处理，有利于识别流程的正常运作，IVR系统能及时收到第四语音识别服务器的识别结果，且用户也能及时得到IVR系统的回应，保证了用户的使用体验。

[0090] 进一步地，本申请一实施例中，上述确定模块，包括：

[0091] 排序单元，用于按照待识别语音数据的总时长从小到大的顺序对所述其它语音识别服务器进行排序；

- [0092] 遍历单元，用于根据所述排序，遍历所述其它语音识别服务器；
- [0093] 判断单元，用于判断当前访问的语音识别服务器是否处于正常工作状态；
- [0094] 确定单元，用于若是，将所述当前访问的语音识别服务器确定为用于处理所述语音识别请求的所述第四语音识别服务器，并停止遍历。
- [0095] 本实施例中，根据上述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器的步骤，具体包括：当第一语音识别服务器为处于无法正常工作状态，则通过对除去第一语音识别服务器之外的其它各语音识别服务器的总时长进行分析，按照待识别语音数据的总时长从小到大的顺序，即递增的方式对其它各语音识别服务器进行排序，并根据该排序遍历上述的其它语音识别服务器，在每遍历到一个语音识别服务器时，如果确定该语音识别服务器处于正常工作状态，则将该语音识别服务器确定为用于处理所述语音识别请求的上述第四语音识别服务器，并将该语音识别请求分发至该语音识别服务器进行处理，并停止遍历。进一步地，如果在遍历结束后，都没有找到处于正常工作状态的语音识别服务器，则会向IVR系统返回处理失败的信息，即当前没有正常可用的语音识别服务器对上述语音识别请求进行处理，另外还会发出维修信息，以使得维修人员对处于无法正常工作状态的语音识别服务器进行维修。
- [0096] 进一步地，本申请一实施例中，上述基于数据分析的资源分配装置，包括：
- [0097] 建立模块，用于与各所述语音识别服务器建立TCP（传输控制协议）长连接，以便通过所述TCP长连接与各所述语音识别服务器之间进行信息交互。
- [0098] 本实施例中，在负载均衡器进行初始化时，可与每个语音识别服务器建立的TCP（Transmission Control Protocol，传输控制协议）长连接，可选地，可与每一个语音识别服务器建立数量至少为1的TCP长连接，其中建立的TCP长连接的具体数量按实际情况进行确定。在该传输控制协议长连接成功建立后，当负载均衡器需要将IVR系统发送的关于某一语音数据的语音识别请求分发至某一语音识别服务器时，即可直接使用上述所建立的传输控制协议长连接，通过所述传输控制协议长连接与该语音识别服务器之间进行信息交互，从而省去了当需要负载均衡器来调用该语音识别服务器时，再去建立传输控制协议长连接所耗费的时间。另外，负载均衡器与每个语音识别服务器之间所建立的传输控制协议长

连接的个数的具体取值可根据实际需要而定，可以为一个，也可以为多个，举例地，负载均衡器与语音识别服务器A建立了一个传输控制协议长连接，负载均衡器与语音识别服务器B建立了两个传输控制协议长连接，负载均衡器与语音识别服务器C建立了五个传输控制协议长连接。其中，建立多个传输控制协议长连接的好处在于：当负载均衡器同时接收到IVR系统发送的多个语音数据对应的多个语音识别请求，且确定多个语音识别请求均需要由同一个语音识别服务器进行处理时，可利用多个传输控制协议长连接分别将多个语音识别请求同时分发至该语音识别服务器，而如果只有一个传输控制协议长连接的话，则只能在分发完一个语音识别请求后再分发另一个语音识别请求，直至将该多个语音识别请求分发完毕，所以建立一定数量的传输控制协议长连接，能够有效的提高了语音识别请求的传输效率。

[0099] 进一步地，本申请一实施例中，上述基于数据分析的资源分配装置，包括：

[0100] 第四判断模块，用于判断是否接收到所述第一语音识别服务器反馈的对于所述语音识别请求的响应信息；

[0101] 转发模块，用于若是，将所述响应信息转发至所述IVR系统；

[0102] 第二接收模块，用于接收由所述IVR系统发送的与所述语音识别请求对应的语音数据；

[0103] 发送模块，用于将所述语音数据发送至所述第一语音识别服务器，以便所述第一语音识别服务器对所述语音数据进行识别。

[0104] 本实施例中，在负载均衡器将所述语音识别请求分发到第一语音识别服务器后，第一语音识别服务器会对该语音识别请求的语法进行解析，如果解析成功则会向负载均衡器返回请求识别成功的响应信息，而如果解析失败，则会向负载均衡器返回错误。在负载均衡器接收到上述响应信息后，会将该响应信息转发至上述的IVR系统，IVR系统在接收到响应信息后会向负载均衡器发送与上述语音识别请求相对应的语音数据，在负载均衡器接收到该语音数据后便会将其转发到该第一语音识别服务器，之后第一语音识别服务器会根据语法，使用对应的语言模型对该语音数据进行识别，得到识别结果，并将该识别结果返回到负载均衡器，使得负载均衡器将该识别结果返回到IVR系统，以实现IVR系统根据

该识别结果与用户进行信息交互。

[0105] 本申请实施例的基于数据分析的资源分配装置，通过对各语音识别服务器当前的待识别语音数据的总时长进行比较分析，来选择合理的语音识别服务器处理接收到的语音识别请求，有利于实现对语音识别服务器的资源均衡使用，提高语音识别请求的处理效率。

[0106] 参照图4，本申请实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图4所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储语音识别请求以及待识别语音数据的时长等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令在执行时，执行如上述各方法的实施例的流程。本领域技术人员可以理解，图4中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定。

[0107] 本申请一实施例还提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，该计算机可读指令在执行时，执行如上述各方法的实施例的流程。

以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

- [权利要求 1] 一种基于数据分析的资源分配方法，其特征在于，包括：
接收IVR（互动式语音应答）系统发送的语音识别请求；
分别计算各语音识别服务器当前的待识别语音数据的总时长；
对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
将所述语音识别请求分发至所述第一语音识别服务器。
- [权利要求 2] 根据权利要求1所述的基于数据分析的资源分配方法，其特征在于，
所述将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：
判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器；
若是，将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。
- [权利要求 3] 根据权利要求2所述的基于数据分析的资源分配方法，其特征在于，
所述将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内的步骤之后，包括：
根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长；
筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器；
经过第一预设时间后，判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值；
若是，将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器，以使得经过分配后的所述第二语音识别服务器内的待识别语音数据的总时长小于所述预设时长阈值。
- [权利要求 4] 根据权利要求1所述的基于数据分析的资源分配方法，其特征在于，

所述将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：

判断所述第一语音识别服务器是否处于正常工作状态；

若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；

根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；

将所述语音识别请求分发至所述第四语音识别服务器。

[权利要求 5]

根据权利要求4所述的基于数据分析的资源分配方法，其特征在于，

所述根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器的步骤，包括：

按照待识别语音数据的总时长从小到大的顺序对所述其它语音识别服务器进行排序；

根据所述排序，遍历所述其它语音识别服务器；

判断当前访问的语音识别服务器是否处于正常工作状态；

若是，将所述当前访问的语音识别服务器确定为用于处理所述语音识别请求的所述第四语音识别服务器，并停止遍历。

[权利要求 6]

根据权利要求1所述的基于数据分析的资源分配方法，其特征在于，

所述接收IVR系统发送的语音识别请求的步骤之前，包括：

与各所述语音识别服务器建立TCP（传输控制协议）长连接，以便通过所述TCP长连接与各所述语音识别服务器进行信息交互。

[权利要求 7]

一种基于数据分析的资源分配装置，其特征在于，包括：

第一接收模块，用于接收IVR系统发送的语音识别请求；

第一计算模块，用于分别计算各语音识别服务器当前的待识别语音数据的总时长；

第一比较模块，用于对各所述总时长进行比较，得出所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；

第一分发模块，用于将所述识别请求分发至所述第一语音识别服务器。

[权利要求 8] 根据权利要求7所述的基于数据分析的资源分配装置，其特征在于，包括：
第一判断模块，用于判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器；
加入模块，用于若是，将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。

[权利要求 9] 根据权利要求8所述的基于数据分析的资源分配装置，其特征在于，包括：
第二计算模块，用于根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长；
筛选模块，用于筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器；
第二判断模块，用于经过第一预设时间后，判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值；
分配模块，用于若是，将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器，以使得经过分配后的所述第二语音识别服务器内的待识别语音数据的总时长小于所述预设时长阈值。

[权利要求 10] 根据权利要求7所述的基于数据分析的资源分配装置，其特征在于，包括：
第三判断模块，用于判断所述第一语音识别服务器是否处于正常工作状态；
第二比较模块，用于若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；

确定模块，用于根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；

第二分发模块，用于将所述语音识别请求分发至所述第四语音识别服务器。

[权利要求 11] 根据权利要求10所述的基于数据分析的资源分配装置，其特征在于，所述确定模块，包括：

排序单元，用于按照待识别语音数据的总时长从小到大的顺序对所述其它语音识别服务器进行排序；

遍历单元，用于根据所述排序，遍历所述其它语音识别服务器；

判断单元，用于判断当前访问的语音识别服务器是否处于正常工作状态；

确定单元，用于若是，将所述当前访问的语音识别服务器确定为用于处理所述语音识别请求的所述第四语音识别服务器，并停止遍历。

[权利要求 12] 根据权利要求7所述的基于数据分析的资源分配装置，其特征在于，包括：

建立模块，用于与各所述语音识别服务器建立TCP（传输控制协议）长连接，以便通过所述TCP长连接与各所述语音识别服务器之间进行信息交互。

[权利要求 13] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现基于数据分析的资源分配方法，该基于数据分析的资源分配方法，包括：

接收IVR（互动式语音应答）系统发送的语音识别请求；

分别计算各语音识别服务器当前的待识别语音数据的总时长；

对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；

将所述语音识别请求分发至所述第一语音识别服务器。

[权利要求 14]

根据权利要求13所述的基于数据分析的计算机设备，其特征在于，所述将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：

判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器；

若是，将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。

[权利要求 15]

根据权利要求14所述的基于数据分析的计算机设备，其特征在于，所述将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内的步骤之后，包括：

根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长；

筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器；

经过第一预设时间后，判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值；

若是，将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器，以使得经过分配后的所述第二语音识别服务器内的待识别语音数据的总时长小于所述预设时长阈值。

[权利要求 16]

根据权利要求13所述的基于数据分析的计算机设备，其特征在于，所述将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：

判断所述第一语音识别服务器是否处于正常工作状态；

若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；

根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；

将所述语音识别请求分发至所述第四语音识别服务器。

- [权利要求 17] 一种计算机非易失性可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现基于数据分析的资源分配方法，该基于数据分析的资源分配方法，包括：
- 接收IVR（互动式语音应答）系统发送的语音识别请求；
- 分别计算各语音识别服务器当前的待识别语音数据的总时长；
- 对各所述总时长进行比较，确定所述待识别语音数据的总时长最少的语音识别服务器，并将所述总时长最少的语音识别服务器确定为用于处理所述语音识别请求的第一语音识别服务器；
- 将所述语音识别请求分发至所述第一语音识别服务器。

- [权利要求 18] 根据权利要求17的计算机非易失性可读存储介质，其特征在于，所述处理器将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：
- 判断是否存在所述待识别语音数据的总时长大于预设时长阈值的第二语音识别服务器；
- 若是，将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内。

- [权利要求 19] 根据权利要求18的计算机非易失性可读存储介质，其特征在于，所述处理器将所述第二语音识别服务器加入禁止处理新的语音识别请求的黑名单内的步骤之后，包括：
- 根据各所述语音识别服务器的待识别语音数据的总时长，计算出各所述语音识别服务器的待识别语音数据的平均时长；
- 筛选出待识别语音数据的总时长小于所述平均时长的第三语音识别服务器；
- 经过第一预设时间后，判断所述第二语音识别服务器内的待识别语音数据的总时长是否大于所述预设时长阈值；
- 若是，将所述第二语音识别服务器内的部分待识别语音数据分配至所述第三语音识别服务器，以使得经过分配后的所述第二语音识别服务

器内的待识别语音数据的总时长小于所述预设时长阈值。

- [权利要求 20] 根据权利要求17的计算机非易失性可读存储介质，其特征在于，所述处理器将所述语音识别请求分发至所述第一语音识别服务器的步骤之后，包括：
- 判断所述第一语音识别服务器是否处于正常工作状态；
- 若否，对除所述第一语音识别服务器之外的其它语音识别服务器当前的待识别语音数据的总时长进行比较，以得到比较结果；
- 根据所述比较结果，确定出用于处理所述语音识别请求的第四语音识别服务器，其中所述第四语音识别服务器处于正常工作状态；
- 将所述语音识别请求分发至所述第四语音识别服务器。

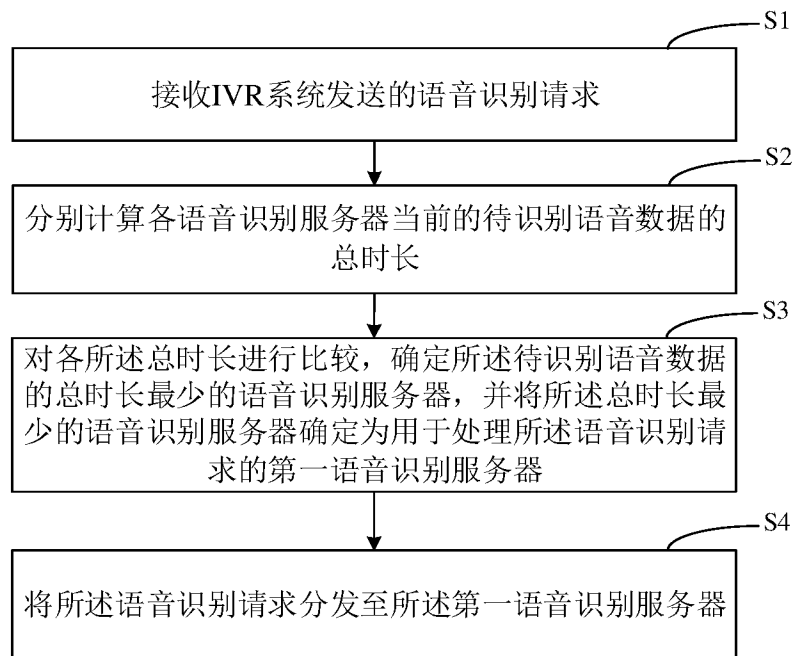


图 1

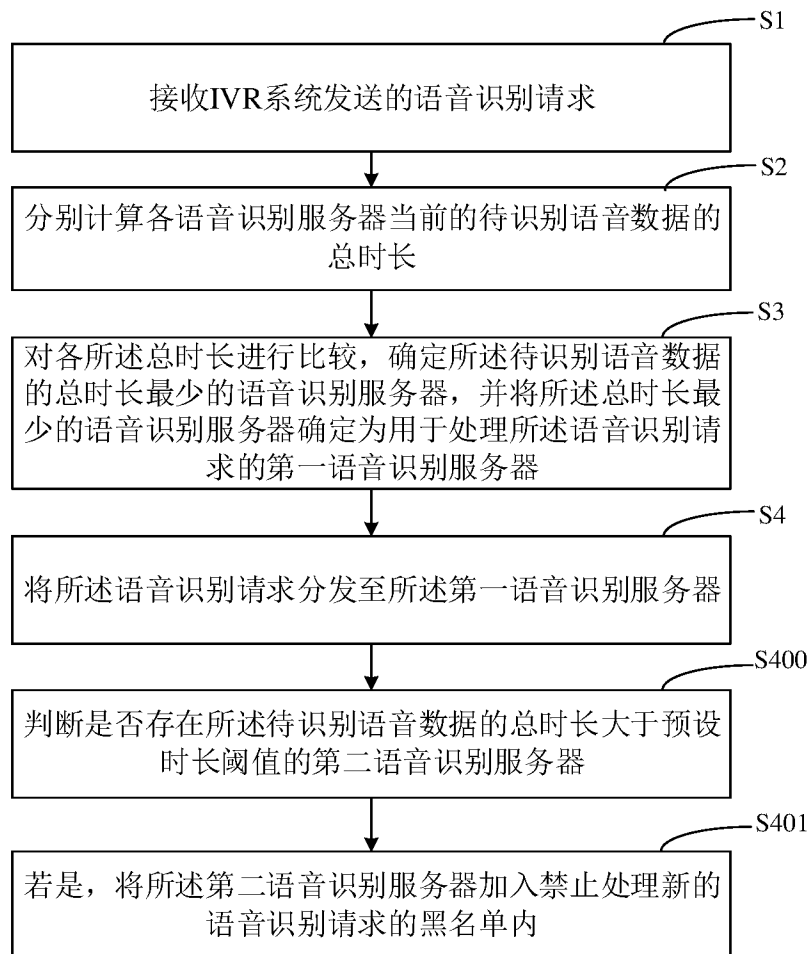


图 2

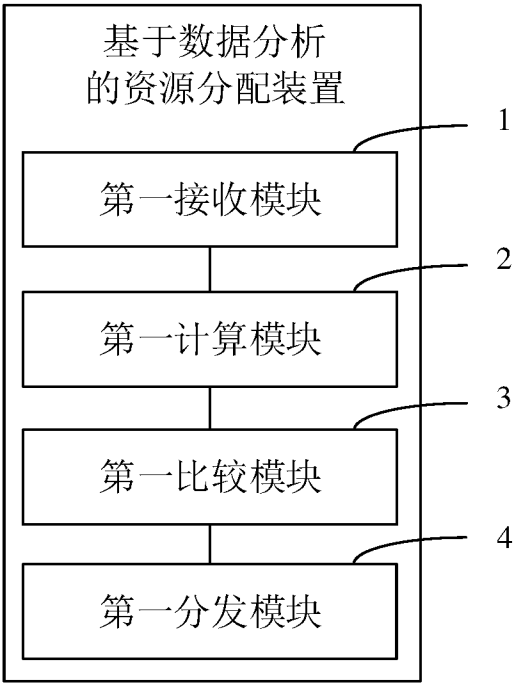


图 3

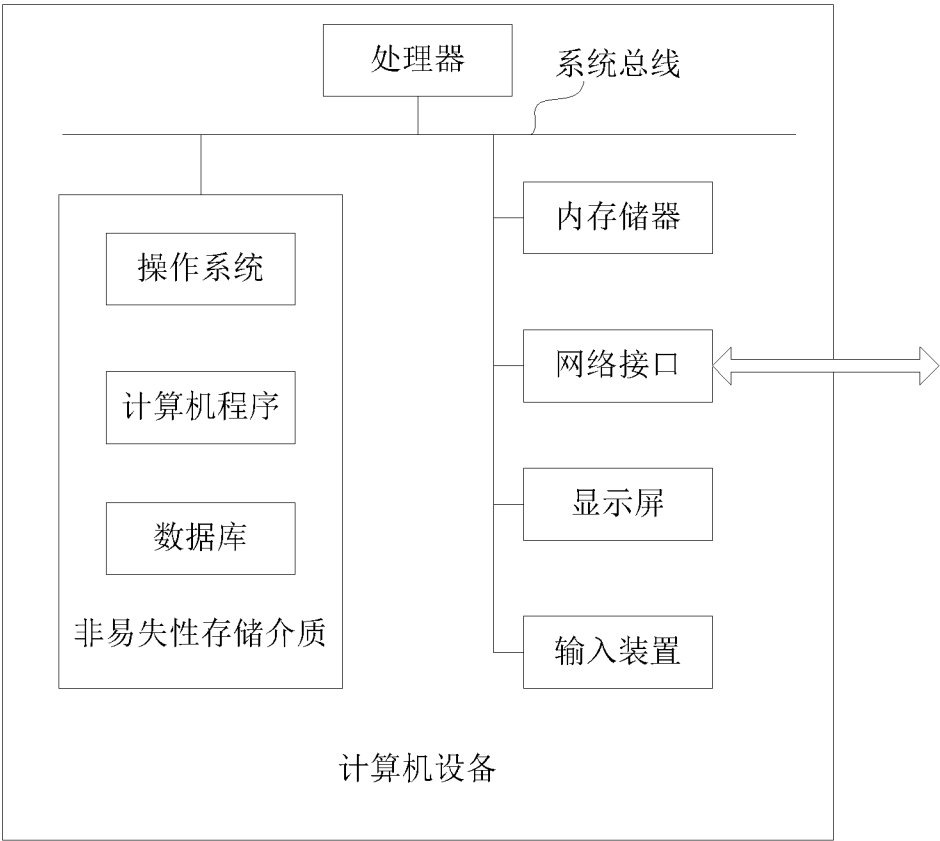


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/124400

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08(2006.01)i; G10L 15/30(2013.01)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)

H04L; G01L

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; WOTXT; USTXT; EPTXT: 待, 服务器, 负载均衡, 最少, 负载均衡, 声音识别, 声音辨识, 语音识别, 语音辨识, 最小, load balance, server, lowest, least, voice recognition, speech recognition

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 103971687 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 06 August 2014 (2014-08-06) description, paragraphs 0008-0075	1-20
Y	CN 108337275 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 27 July 2018 (2018-07-27) description, paragraph 0019	1-20
Y	CN 106101232 A (BEIJING SIYUAN LANDMARK TECHNOLOGY CO., LTD.) 09 November 2016 (2016-11-09) description, paragraphs 0005-0047	1-20
A	CN 1571317 A (HUAWEI TECHNOLOGIES CO., LTD.) 26 January 2005 (2005-01-26) entire document	1-20
A	CN 106847291 A (CHENGDU CHIPINTELLI CO., LTD.) 13 June 2017 (2017-06-13) 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

“D” document cited by the applicant in the international application

“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

24 July 2019

Date of mailing of the international search report

31 July 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (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/CN2018/124400

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	103971687	A	06 August 2014	CN	103971687 B	29 June 2016
				WO	2014117584 A1	07 August 2014
				CA	2898783 A1	07 August 2014
				JP	2016507079 A	07 March 2016
				US	2014337022 A1	13 November 2014
				JP	5951148 B2	13 July 2016
				SG	11201505611 A1	28 August 2015
				SG	11201505611 B	16 May 2016
CN	108337275	A	27 July 2018	None		
CN	106101232	A	09 November 2016	None		
CN	1571317	A	26 January 2005	None		
CN	106847291	A	13 June 2017	None		

国际检索报告

国际申请号

PCT/CN2018/124400

A. 主题的分类

H04L 29/08(2006.01)i; G10L 15/30(2013.01)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)

H04L; G01L

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; WOTXT; USTXT; EPTXT: 待, 服务器, 负载平衡, 最少, 负载均衡, 声音识别, 声音辨识, 语音识别, 语音辨识, 最小, load balance, server, lowest, least, voice recognition, speech recognition

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
Y	CN 103971687 A (腾讯科技深圳有限公司) 2014年 8月 6日 (2014 - 08 - 06) 说明书第0008-0075段	1-20
Y	CN 108337275 A (百度在线网络技术北京有限公司) 2018年 7月 27日 (2018 - 07 - 27) 说明书第0019段	1-20
Y	CN 106101232 A (北京思源置地科技有限公司) 2016年 11月 9日 (2016 - 11 - 09) 说明书第0005-0047段	1-20
A	CN 1571317 A (华为技术有限公司) 2005年 1月 26日 (2005 - 01 - 26) 全文	1-20
A	CN 106847291 A (成都启英泰伦科技有限公司) 2017年 6月 13日 (2017 - 06 - 13) 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

2019年 7月 24日

国际检索报告邮寄日期

2019年 7月 31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

罗芳洁

电话号码 86-(010)-62411638

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/124400

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	103971687	A	2014年 8月 6日	CN 103971687 B	2016年 6月 29日
				WO 2014117584 A1	2014年 8月 7日
				CA 2898783 A1	2014年 8月 7日
				JP 2016507079 A	2016年 3月 7日
				US 2014337022 A1	2014年 11月 13日
				JP 5951148 B2	2016年 7月 13日
				SG 11201505611 A1	2015年 8月 28日
				SG 11201505611 B	2016年 5月 16日
CN	108337275	A	2018年 7月 27日	无	
CN	106101232	A	2016年 11月 9日	无	
CN	1571317	A	2005年 1月 26日	无	
CN	106847291	A	2017年 6月 13日	无	