(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113094559 A
(43)申请公布日 2021.07.09

(21)申请号 202110450705.3

(22)申请日 2021.04.25

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 赵志新 庞敏辉 肖岩

(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 范芳茗

(51)Int.Cl.
G06F 16/903(2019.01)
G06F 16/9032(2019.01)

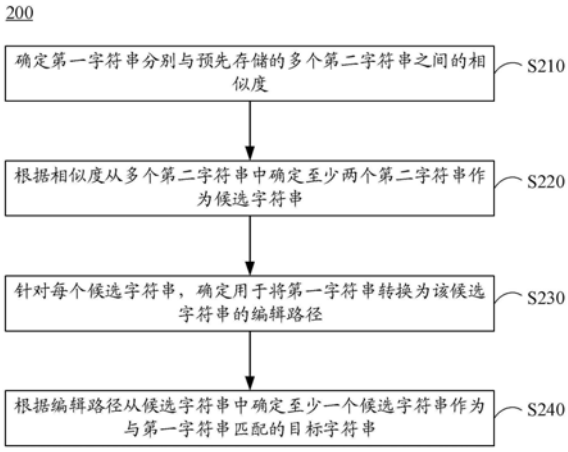
权利要求书5页 说明书12页 附图5页

(54)发明名称

信息匹配方法、装置、电子设备和存储介质

(57)摘要

本公开提供了一种信息匹配方法，涉及人工智能领域，尤其涉及语音识别领域。具体实现方案为：确定第一字符串分别与预先存储的多个第二字符串之间的相似度；根据相似度从多个第二字符串中确定至少两个第二字符串作为候选字符串；针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径；根据编辑路径从候选字符串中确定至少一个候选字符串作为与第一字符串匹配的目标字符串。本公开还公开了一种信息匹配装置、电子设备和存储介质。



CN 113094559 A

1.一种信息匹配方法,包括:

确定第一字符串分别与预先存储的多个第二字符串之间的相似度;

根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串;

针对每个候选字符串,确定用于将第一字符串转换为该候选字符串的编辑路径;

根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串。

2.根据权利要求1所述的方法,其中,所述确定第一字符串分别与预先存储的多个第二字符串之间的相似度包括:

将第一字符串转换为字符序列;

确定所述字符序列分别与多个索引序列之间的相似度作为所述第一字符串分别与多个第二字符串之间的相似度,其中,所述多个索引序列分别是由所述多个第二字符串转换得到的。

3.根据权利要求2所述的方法,其中,所述将第一字符串转换为字符序列包括:

将第一字符串按照第一规则切分为多个第一字符片段;

将第一字符串按照第二规则切分为多个第二字符片段;

在所述多个第一字符片段中的至少一个字符片段中,添加该第一字符片段在第一字符串中的位置信息,得到第一字符片段的序列;

在所述多个第二字符片段中的至少两个字符片段中,添加该第二字符片段在第一字符串中的位置信息,得到第二字符片段的序列;

基于第一字符片段的序列和第二字符片段的序列,生成所述字符序列。

4.根据权利要求3所述的方法,其中,所述第一规则包括以单个字符作为一个切分单元进行切分,所述第二规则包括以任意两个相邻字符作为切分单元进行切分。

5.根据权利要求1所述的方法,其中,

在多个索引序列中不存在与字符序列相同的索引序列的情况下,执行所述根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串的操作。

6.根据权利要求1所述的方法,其中,所述确定用于将第一字符串转换为候选字符串的编辑路径包括:

计算第一字符串与候选字符串之间的最短编辑距离,所述最短编辑距离表示将第一字符串转换为候选字符串的最少操作次数;

确定在满足最短编辑距离的情况下将第一字符串转换为候选字符串的编辑操作集合,作为编辑路径。

7.根据权利要求1所述的方法,其中,所述根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串包括:

根据每个候选字符串的编辑路径中包含的编辑操作的类型和数量,对所述编辑路径进行打分,并根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度;

从所述候选字符串中确定具有符合预设条件的匹配度的至少一个候选字符串,作为与所述第一字符串匹配的目标字符串。

8.根据权利要求7所述的方法,其中,所述根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度包括:

在候选字符串存在多个编辑路径的情况下，根据多个编辑路径的得分中的最高分来计算该候选字符串与第一字符串之间的匹配度。

9.根据权利要求7所述的方法，其中，所述对所述编辑路径进行打分包括：

根据所述编辑路径中包含的编辑操作的类型和数量来确定所述编辑路径的扣分分数；

根据针对所述编辑路径而预先设置的基础分数和所述编辑路径的扣分分数，确定所述编辑路径的得分。

10.根据权利要求9所述的方法，其中，所述第一字符串是从语音输入信息中识别得到的，所述确定所述编辑路径的扣分分数包括：

针对编辑路径中包含的每个编辑操作，

在所述编辑操作为删除操作的情况下，将所述删除操作的扣分分数确定为第一预设值；

在所述编辑操作为增加操作的情况下，将所述增加操作的扣分分数确定为第二预设值；

在所述编辑操作为替换操作的情况下，根据第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度，确定所述替换操作的扣分分数，以及

基于编辑路径中包含的至少一个编辑操作的扣分分数，来计算所述编辑路径的扣分分数，所述至少一个编辑操作包括所述删除操作、所述增加操作和/或所述替换操作。

11.根据权利要求10所述的方法，其中，所述根据第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度，确定所述替换操作的扣分分数包括：

计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度；

根据以下公式计算所述编辑路径中的替换操作的扣分分数：

$q = -A*(1-p)$

其中，q表示所述编辑路径中的替换操作的扣分分数，A为预设参数，p表示第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

12.根据权利要求11所述的方法，其中，所述计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度包括：

基于预设的拼音编码映射表，生成第一字符串中被替换字符的第一拼音编码以及候选字符串中替换字符的第二拼音编码；

计算第一拼音编码和第二拼音编码之间的最短编辑距离；

根据第一拼音编码和第二拼音编码之间的最短编辑距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

13.根据权利要求11所述的方法，其中，所述计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度包括：

使用语音识别模型获取第一字符串中被替换字符的第一语音向量和候选字符串中替换字符的第二语音向量；

基于第一语音向量和第二语音向量之间的距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

14.根据权利要求1至13中任一项权利要求所述的方法，其中，所述字符串为车牌号。

15.一种信息匹配装置，包括：

第一确定模块，用于确定第一字符串分别与预先存储的多个第二字符串之间的相似度；

第二确定模块，用于根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串；

第三确定模块，用于针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径；

第四确定模块，用于根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串。

16.根据权利要求15所述的装置，其中，所述第一确定模块包括：

转换单元，用于将第一字符串转换为字符序列；

第一确定单元，确定所述字符序列分别与多个索引序列之间的相似度作为所述第一字符串分别与多个第二字符串之间的相似度，其中所述多个索引序列分别是由所述多个第二字符串转换得到的。

17.根据权利要求15所述的装置，其中，所述串转换单元包括：

第一切分子单元，将第一字符串按照第一规则切分为多个第一字符片段；

第二切分子单元，将第一字符串按照第二规则切分为多个第二字符片段；

第一添加子单元，在所述多个第一字符片段中的至少一个字符片段中，添加该第一字符片段在第一字符串中的位置信息，得到第一字符片段的序列；

第二添加子单元，在所述多个第二字符片段中的至少二个字符片段中，添加该第二字符片段在第一字符串中的位置信息，得到第二字符片段的序列；

字符序列生成子单元，基于第一字符片段的序列和第二字符片段的序列，生成所述字符序列。

18.根据权利要求17所述的装置，其中，所述第一规则包括以单个字符作为一个切分单元进行切分，所述第二规则包括以任意两个相邻字符作为切分单元进行切分。

19.根据权利要求15所述的装置，其中，所述第二确定模块被配置为在多个索引序列中不存在与字符序列相同的索引序列的情况下，执行所述根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串的操作。

20.根据权利要求15所述的装置，其中，所述第三确定模块包括：

第一计算单元，用于计算第一字符串与候选字符串之间的最短编辑距离，所述最短编辑距离表示将第一字符串转换为候选字符串的最少操作次数；

第二确定单元，用于确定在满足最短编辑距离的情况下将第一字符串转换为候选字符串的编辑操作集合，作为编辑路径。

21.根据权利要求15所述的装置，其中，所述第四确定模块包括：

打分单元，用于根据每个候选字符串的编辑路径中包含的编辑操作的类型和数量，对所述编辑路径进行打分；

第二计算单元，用于根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度；

第三确定单元，用于从所述候选字符串中确定具有符合预设条件的匹配度的至少一个候选字符串，作为与所述第一字符串匹配的目标字符串。

22.根据权利要求21所述的装置,其中,所述第二计算单元用于在候选字符串存在多个编辑路径的情况下,根据多个编辑路径的得分中的最高分来计算该候选字符串与第一字符串之间的匹配度。

23.根据权利要求21所述的装置,其中,所述打分单元包括:

第一确定子单元,用于根据所述编辑路径中包含的编辑操作的类型和数量来确定所述编辑路径的扣分分数;

第二确定子单元,根据针对所述编辑路径而预先设置的基础分数和所述编辑路径的扣分分数,确定所述编辑路径的得分。

24.根据权利要求23所述的装置,其中,所述第一字符串是从语音输入信息中识别得到的,所述第一确定子单元包括:

扣分确定子单元,用于针对编辑路径中包含的每个编辑操作,在所述编辑操作为删除操作的情况下,将所述删除操作的扣分分数确定为第一预设值;在所述编辑操作为增加操作的情况下,将所述增加操作的扣分分数确定为第二预设值;在所述编辑操作为替换操作的情况下,根据第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度,确定所述替换操作的扣分分数,以及

扣分计算子单元,基于编辑路径中包含的至少一个编辑操作的扣分分数,来计算所述编辑路径的扣分分数。

25.根据权利要求24所述的装置,其中,所述扣分确定子单元包括:

发音相似度计算子单元,用于计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度;

替换操作扣分子单元,用于根据以下等式计算所述编辑路径中的替换操作的扣分分数:

$$q=-A*(1-p)$$

其中,q表示所述编辑路径中的替换操作的扣分分数,A为预设参数,p表示第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

26.根据权利要求25所述的装置,其中,所述发音相似度计算子单元包括:

拼音映射子单元,用于基于预设的拼音编码映射表,生成第一字符串中被替换字符的第一拼音编码以及候选字符串中替换字符的第二拼音编码;

距离计算子单元,用于计算第一拼音编码和第二拼音编码之间的最短编辑距离;

第一发音相似度确定子单元,用于根据第一拼音编码和第二拼音编码之间的最短编辑距离,计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

27.根据权利要求25所述的装置,其中,所述发音相似度计算子单元包括:

语音识别子单元,用于使用语音识别模型获取第一字符串中被替换字符的第一语音向量和候选字符串中替换字符的第二语音向量;

第二发音相似度确定子单元,用于基于第一语音向量和第二语音向量之间的距离,计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

28.根据权利要求15至27中任一项权利要求所述的装置,其中,所述字符串为车牌号。

29.一种电子设备,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权利要求1至14中任一项所述的方法。

30.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1至14中任一项所述的方法。

31.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1至14中任一项所述的方法。

信息匹配方法、装置、电子设备和存储介质

技术领域

[0001] 本公开涉及人工智能技术领域，尤其涉及语音识别技术。更具体地，本公开提供了一种信息匹配方法、装置、电子设备和存储介质。

背景技术

[0002] 随着互联网的快速发展，智能客服类产品与日俱增，需要能够准确地根据用户输入的字符串来检索到匹配结果。例如在汽车行业的114挪车服务和车险客服服务等，需要根据用户输入的出车牌号在系统数据库中查询是否有对应的车牌号，以便进行车辆服务。然而在实践中常常由于各种原因导致用户输入的车牌号的信息不准确，从而导致无法获得准确的匹配结果。

发明内容

[0003] 本公开提供了一种信息匹配方法、装置、设备以及存储介质。

[0004] 根据第一方面，提供了一种信息匹配方法，该方法包括：确定第一字符串分别与预先存储的多个第二字符串之间的相似度；根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串；针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径；根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串。

[0005] 根据第二方面，提供了一种信息匹配装置，该装置包括：第一确定模块，用于确定第一字符串分别与预先存储的多个第二字符串之间的相似度；第二确定模块，用于根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串；第三确定模块，用于针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径；第四确定模块，用于根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串。

[0006] 根据第三方面，提供了一种电子设备，包括：至少一个处理器；以及与至少一个处理器通信连接的存储器；其中，存储器存储有可被至少一个处理器执行的指令，指令被至少一个处理器执行，以使至少一个处理器能够执行根据本公开提供的方法。

[0007] 根据第四方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，该计算机指令用于使计算机执行根据本公开提供的方法。

[0008] 根据第五方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现根据本公开提供的方法。

[0009] 应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0010] 附图用于更好地理解本方案，不构成对本公开的限定。其中：

[0011] 图1是根据本公开的一个实施例的可以应用信息匹配方法和装置的示例性系统架构示意图；

[0012] 图2是根据本公开的一个实施例的信息匹配方法的流程图；

[0013] 图3是根据本公开的一个实施例的确定用于将第一字符串转换为候选字符串的编辑路径的方法的流程图；

[0014] 图4是根据本公开的一个实施例的确定至少一个候选字符串作为与第一字符串匹配的目标字符串的方法的流程示意图；

[0015] 图5是根据本公开的一个实施例的计算发音相似度的方法的流程示意图；

[0016] 图6是根据本公开的另一个实施例的计算发音相似度的方法的流程示意图；

[0017] 图7是根据本公开的另一个实施例的信息匹配方法的流程示意图；

[0018] 图8是根据本公开的一个实施例的信息装置的框图；

[0019] 图9是根据本公开的一个实施例的信息匹配方法的电子设备的框图。

具体实施方式

[0020] 以下结合附图对本公开的示范性实施例做出说明，其中包括本公开实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本公开的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

[0021] 在智能客服场景下，有些业务需要用户输入车牌号码，如114挪车服务、车险客服服务等。电话或语音客服系统中，需要对用户的语音输入做ASR(Automatic Speech Recognition，自动语音识别技术)识别，将语音输入信息中的车牌号转为文本车牌号之后，在系统数据库中查询是否有对应的车牌号，如果数据库中不存在，则无法继续办理业务。在语音ASR过程，由于噪音、用户方言等原因，非常容易出现识别错误，为了保证服务能够正常继续，需要对车牌数据库匹配不到的用户输入的车牌号码，与车牌数据库中车牌号码做模糊匹配，在误差允许范围内，选择最接近的车牌号码，然后可根据需求向用户二次询问，或直接使用模糊匹配到的车牌号码办理后续业务。

[0022] 图1是根据本公开一个实施例的可以应用信息匹配方法和装置的示例性系统架构示意图。需要注意的是，图1所示仅为可以应用本公开实施例的系统架构的示例，以帮助本领域技术人员理解本公开的技术内容，但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0023] 如图1所示，根据该实施例的系统架构100可以包括终端设备101、网络102和服务器103。网络102用以在终端设备101和服务器103之间提供通信链路的介质。网络102可以包括各种连接类型，例如有线和/或无线通信链路等等。

[0024] 终端设备101可以是智能终端，例如手机、平板电脑等。终端设备101上可以安装有智能客服服务的客户端。用户在使用智能客服服务的场景中，可以通过终端设备101输入字符串，输入形式可以包括语音输入、图像输入和文本输入等，终端设备101将输入信息通过网络102发送给服务器103，服务器103从输入信息中识别出输入字符串，然后使用输入字符串与字符串库中的字符串进行匹配，如果从字符串库中匹配到与输入字符串相同的字符串，则可以进行业务办理等服务。如果没有从字符串库中匹配到与输入字符串相同的字符

串,可以基于输入字符串与字符串库中的字符串之间的相似度来获取多个候选字符串,从候选字符串中确定与输入字符串之间相似度最高的至少一个字符串作为与输入字符串相匹配的字符串,并返回给终端设备101,供用户选择。

[0025] 在一些实施例中,字符串可以是车牌号。然而本公开不限于此,字符串可以为任何其他形式的字符串,包括但不限于按照预设规则编辑并且长度在预设范围内的字符串,例如ID号码、数据库中的文档编号等等。

[0026] 图2是根据本公开的一个实施例的信息匹配方法的流程图。

[0027] 如图2所示,该信息匹配方法200可以包括操作S210～操作S240。

[0028] 在操作S210,确定第一字符串分别与预先存储的多个第二字符串之间的相似度。

[0029] 例如,第一字符串可以包含在用户输入的语音信息、图像信息或文本信息中,从语音信息、图像信息或文本信息中可以识别出第一字符串。在从语音信息中识别出第一字符串的过程中,需要考虑到语音输入的信息准确性。例如在字符串是车牌号的情况下,用户的语音当中不一定只包含车牌号码,可能带有一些其他的口语信息,如“车牌是”、“我看看”等等。此时,需要对识别出的内容做预处理,例如,可以将识别出的内容的头部及尾部进行删除操作,得到准确的第一字符串。

[0030] 多个第二字符串例如可以预先存储在字符串库中。第一字符串分别与预先存储的多个第二字符串之间的相似度可以基于文本相似度、最短编辑距离或Jaccard相似度来计算得到。

[0031] 在字符串库中还可以创建各个第二字符串的索引数据,例如ES(Elastic Search)索引,在使用第一字符串与多个第二字符串进行匹配时,可以将第一字符串也转换为ES索引的格式,通过计算第一字符串的索引数据与字符串库中第二字符串的索引数据之间的相似度来确定第一字符串与第二字符串之间的相似度。

[0032] 在操作S220,根据相似度从多个第二字符串中确定至少两个第二字符串作为候选字符串。

[0033] 具体地,可以根据相似度从多个第二字符串中确定至少两个候选字符串。例如,可以获取与第一字符串之间的相似度高于预设阈值(例如80%)的第二字符串作为候选字符串。又例如,按照与第一字符串之间的相似度高进行降序排列,从排序中选取预设数量(例如前10个或15个等)的第二字符串作为候选字符串。

[0034] 在操作S230,针对每个候选字符串,确定用于将第一字符串转换为该候选字符串的编辑路径。

[0035] 在确定候选字符串之后,可以基于编辑路径从候选字符串中确定与第一字符串相匹配的目标字符串,具体可以确定用于将第一字符串转换为各个候选字符串的编辑路径,并对编辑路径进行打分,根据编辑路径的得分来确定目标字符串。

[0036] 针对每个候选字符串,可以通过至少一个编辑操作将第一字符串转换为该候选字符串,则将第一字符串转换为候选字符串的编辑路径可以是这些编辑操作的集合。

[0037] 例如,在字符串为车牌号的情况下,假设第一车牌号是“黑LB4099”,候选车牌号是“黑LD409”。将第一车牌号转换为候选车牌号的编辑操作可以包括“操作1,将第三位的B改写为D;操作2,删除末位9”,则操作1和操作2组成一个用于将第一车牌号转换为候选车牌号的编辑路径(可以简称为第二车牌号的编辑路径)。

[0038] 在操作S240,根据编辑路径从候选字符串中确定至少一个候选字符串作为与第一字符串匹配的目标字符串。

[0039] 例如,可以根据各个候选字符串的编辑路径中的编辑操作类型和数量,来对各个候选字符串进行打分,根据各个候选字符串的得分来确定与第一字符串匹配的候选字符串作为目标字符串。目标字符串可以有一个或多个,可以给用户更多选择。

[0040] 具体地,候选字符串的编辑路径中包含了至少一个编辑操作,编辑操作的类型包括增加、删除和替换,可以预先设置编辑路径的基础分数,并确定每个类型的编辑操作的扣分分数,根据编辑路径中的至少一个编辑操作的扣分分数来计算编辑路径的扣分分数,根据针对编辑路径而预先设置的基础分数和编辑路径的扣分分数,确定编辑路径的得分。

[0041] 编辑操作的扣分分数可以表示基于该编辑操作将第一字符串转换为候选字符串的代价,如果将第一字符串转换为候选字符串需要进行替换操作,则可能是因为用户口音、环境噪音等因素导致的语音识别出错,如用户语音输入为“1”,语音识别为“7”,这种情况应给予替换操作较小的扣分分数(如-0.05)。增加和删除操作可以分配较大的扣分分数(如-0.1)。

[0042] 候选字符串的编辑路径的得分可以作为候选字符串的得分来衡量第一字符串与该候选字符串之间的匹配度。根据第一字符串与各个候选字符串之间的匹配度从候选字符串中一个或多个目标字符串。

[0043] 根据本公开的实施例,相比于直接根据第一字符串与字符串库中的第二字符串之间的相似度来确定与第一字符串相匹配的目标字符串,本公开实施例基于相似度从字符串库中确定至少两个候选字符串,基于编辑路径对至少两个候选字符串进一步进行精细排序,从至少两个候选字符串中确定出与第一字符串相匹配的目标字符串,能够提字符串的匹配准确率。

[0044] 在字符串库中可以创建各个第二字符串的弹性检索(ES,ElasticSearch)索引序列,通过ES检索来获得与第一字符串相似的候选字符串。ES是一种分布式全文检索方法,其基本原理是对提交到数据库中的数据进行切分并将切分结果与权重一并存储,在检索时根据权重对结果排名或打分,根据排名或打分结果来返回检索结果。这里第二字符串的索引序列可以指的是通过对第二字符串进行切分得到的多个字符片段的序列。由于字符串库数量级庞大,需要扩大召回数量保证ES召回率,为了提高检索效率和准确率,可以结合第二字符串中字符的位置信息来创建ES索引序列,利用字符位置来提高检索效率和准确率。具体可以基于预设的切分规则将第二字符串切分为多个字符片段,在字符片段中添加该字符片段在第二字符串中的位置信息。

[0045] 预设的切分规则可以包括第一切分规则和第二切分规则。第一切分规则可以包括将第二字符串按照单字符作为切分单元来进行切分,第二切分规则可以包括将第二字符串按照相邻两个字符作为切分单元来进行切分。在一些实施例中,第一切分规则可以包括Uni-Gram切分规则,第二切分规则可以包括Bi-Gram切分规则。但是本公开的实施例不限于此,可以根据需要选择其他合适的切分规则来进行切分。添加位置信息可以是在多个字符片段中对第二字符串首部预设个数(例如2个)的字符或尾部预设个数的字符分别添加开始标识(例如“S”)和结束的标识(例如“E”)。

[0046] 例如,在字符串为车牌号的情况下,针对第二车牌号“黑LB4099”,基于Uni-Gram切

分规则可以得到多个第一字符片段的组合[″黑″,″L″,″B″,″4″,″0″,″9″,″9″],基于Bi-Gram切分规则可以得到多个第二字符片段的组合[″黑L″,″LB″,″B4″,″40″,″09″,″99″]。在多个第一字符片段的组合中添加位置信息,例如用“S”来表示位于起始两位的字符片段,用“E”来表示位于最后两位的字符片段,从而可以得到[″黑-S″,″L-S″,″B″,″4″,″0″,″9-E″,″9-E″],在多个第一字符片段的组合中添加位置信息可以得到[″黑L-S″,″LB-S″,″B4″,″40″,″09-E″,″99-E″],则第二车牌号的完整的索引序列可以是[″黑-S″,″L-S″,″B″,″4″,″0″,″9-E″,″9-E″,″黑L-S″,″LB-S″,″B4″,″40″,″09-E″,″99-E″]。

[0047] 针对用户输入的第一字符串采用上述同样的方式创建第一字符串的字符序列,使用字符序列对车牌库中的索引序列进行检索,由于在字符序列和索引序列中均添加了第一字符串和第二字符串的字符的位置信息,在进行检索时能够高效利用第一字符串和第二字符串中每个字符的位置,提高检索效率和准确率。所谓第一字符串的“字符序列”可以是指通过对第一字符串进行切分得到的多个字符片段的序列。例如在字符串为车牌号的情况下,对于车牌号“京NKM319”,则通过按照与第二车牌号相同的方式进行处理,可以得到字符序列[″京-S″,″N-S″,″K″,″M″,″3″,″1-E″,″5-E″,″黑N-S″,″NK-S″,″KM″,″M3″,″31-E″,″19-E″]。

[0048] 本公开的实施例在创建ES索引序列中使用了Uni-Gram和Bi-Gram的切分规则,在仅有1位字符出错时,利用Bi-Gram相邻字符的前后关系查询更为精确,在车牌号码发生2位错误时,依靠Uni-Gram可以保证召回数量。根据实际需求,还可以使用更多的切分规则,例如,按照相邻三个字符进行切分的规则和按照相邻四个字符进行切分的规则等等。

[0049] 图3是根据本公开的一个实施例的确定用于将第一字符串转换为候选字符串的编辑路径的方法的流程图。

[0050] 如图3所示,该方法可以包括操作S331～操作S332。

[0051] 在操作S331,计算第一字符串与候选字符串之间的最短编辑距离。其中,最短编辑距离表示将第一字符串转换为候选字符串的最少操作次数。

[0052] 例如,针对第一字符串和候选字符串,可以使用动态规划算法计算第一字符串与候选字符串之间的最短编辑距离。动态规划算法可以是通过遍历将第一字符串转换为候选字符串的所有操作方法,每种操作方法包括至少一个编辑操作。确定所有操作方法中使用的编辑操作的次数的最小值作为第一字符串与候选字符串之间的最短编辑距离。

[0053] 例如,将第一字符串A转换为候选字符串B可以有十种操作方法,十种操作方法中使用编辑操作的次数最小为2,则第一字符串A与候选字符串B之间的最小编辑距离为2。

[0054] 在操作S332,确定在满足最短编辑距离的情况下将第一字符串转换为候选字符串的编辑操作集合,作为编辑路径。

[0055] 例如,可以通过回溯算法从遍历的所有操作方法中确定满足最短编辑距离的操作方法,该操作方法包含的编辑操作的集合为一个编辑路径。

[0056] 满足最短距离的操作方法可以是一个或多个,则将第一字符串转换为候选字符串的编辑路径对应地也是一个或多个。在编辑路径是多个的情况下,该多个编辑路径中每个编辑路径所包含的编辑操作次数相同,但是每个编辑路径所包含的编辑操作不完全相同,具体可以是编辑操作的类型不同,或者编辑操作的位置不同。

[0057] 例如,在字符串为车牌号的情况下将第一车牌号“白L83”转换为候选车牌号“白

L8333”的最短编辑距离是2,但是满足该最短编辑距离的编辑路径可以有多个,其中一个编辑路径包括“操作1,在末位添加3;操作2,在末位添加3”,另一个编辑路径可以是“操作1,在第4位添加3;操作2,在末位添加3”,还有其他的编辑路径这里不再举例。

[0058] 根据本公开的实施例,在获得候选字符串之后,基于候选字符串的编辑路径对候选字符串进行打分,能够对候选字符串进行进一步地精细排序,提高字符串的匹配准确率。

[0059] 图4是根据本公开的一个实施例的确定至少一个候选字符串作为与第一字符串匹配的目标字符串的方法的流程图。

[0060] 如图4所示,该方法可以包括操作S441～操作S442。

[0061] 在操作S441,根据每个候选字符串的编辑路径中包含的编辑操作的类型和数量,对编辑路径进行打分,并根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度。

[0062] 例如,编辑操作的类型包括增加、删除和替换。在对每个候选字符串的编辑路径进行打分时,可以预先设置编辑路径的基础分数,并确定不同类型的编辑操作的扣分分数,根据编辑路径中所包含的编辑操作的扣分分数来确定编辑路径的扣分分数,根据针对编辑路径而预先设置的基础分数和编辑路径的扣分分数,确定编辑路径的得分。其中,编辑操作的扣分分数可以表示基于该编辑操作将第一字符串转换为候选字符串的代价。

[0063] 例如,针对删除操作和增加操作,由于语音识别过程中出现少识别一位符号或多识别一位符号的概率较小,如果将第一字符串转换为候选字符串需要进行删除操作或增加操作,则大概率是由于该候选字符串与第一字符串不匹配,因此可以给予删除操作和增加操作较大的扣分分数,例如对删除操作和增加操作均设置-0.1的扣分分数。

[0064] 针对替换操作,由于用户口音或环境噪音的影响,在语音识别过程中可能出现将输入字符识别为发音相近的其他字符(例如,将1识别为7,将B识别为D等等),这种情况下,应给予替换操作较小的扣分。因此,如果第一字符串转换为候选字符串需要进行替换操作,应根据第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度来确定替换操作的扣分分数。

[0065] 如果候选字符串的编辑距离有多个,则可以以多个编辑距离的得分中的最高分为准。在确定编辑路径的得分之后,可以将候选字符串的编辑路径的得分可以作为候选字符串的得分来衡量第一字符串与该候选字符串之间的匹配度。

[0066] 在操作S442,从候选字符串中确定具有符合预设条件的匹配度的至少一个候选字符串,作为与第一字符串匹配的目标字符串。

[0067] 根据第一字符串与各个候选字符串之间的匹配度从候选字符串中一个或多个目标字符串。例如,候选字符串的得分作为该候选字符串与第一字符串之间的匹配度,可以将得分大于预设阈值(例如0.8)的候选字符串作为匹配成功的目标字符串,目标字符串如果有多个,可以首先将与第一字符串匹配度最高的目标字符串反馈给用户,询问用户输入的第一字符串是否为该目标字符串,如果用户选择否,则可以使用与第一字符串之间的匹配度次高的目标字符串向用户进行询问,直至用户确定正确的目标字符串,多个目标字符串可以给用户更多选择。

[0068] 根据本公开的实施例,在获得候选字符串之后,基于候选字符串的编辑路径对候选字符串进行打分,能够对候选字符串进行进一步地精细排序,提高字符串的匹配准确率。

[0069] 在确定编辑路径中的替换操作的扣分分数时，可以根据以下等式计算编辑路径中的替换操作的扣分分数：

[0070] q＝-A*(1-p)

[0071] 其中，q表示编辑路径中的替换操作的扣分分数，A为预设参数(A可以是0.05)，p表示第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0072] 发音相似度可以基于替换前后的两个字符的拼音编辑距离来计算，还可以基于替换前后的两个字符的语音向量之间的距离来计算。

[0073] 图5是根据本公开的一个实施例的计算发音相似度的方法的流程图。

[0074] 如图5所示，该方法可以包括操作S5411～操作S5413。

[0075] 在操作S5411，基于预设的拼音编码映射表，生成第一字符串中被替换字符的第一拼音编码以及候选字符串中替换字符的第二拼音编码。

[0076] 在操作S5412，计算第一拼音编码和第二拼音编码之间的最短编辑距离。

[0077] 在操作S5413，根据第一拼音编码和第二拼音编码之间的最短编辑距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0078] 例如，可以使用预设的拼音编码映射表将对替换前后两个字符的拼音进行编码，分别得到第一拼音编码和第二拼音编码，计算第一拼音编码和第二拼音编码之间的最短编辑距离，确定第一拼音编码和第二拼音编码中的最大长度，计算第一拼音编码和第二拼音编码之间的最短编辑距离与两个拼音编码中的最大长度之间的比值，该比值可以作为第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0079] 根据本公开的实施例，基于第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度来对候选字符串进行打分，考虑到了在语音识别过程中发音相似的字符被识别错误的情况，能够更精确地对候选字符串进行打分，提高候选字符串的匹配准确率。

[0080] 图6是根据本公开的另一个实施例的计算发音相似度的方法的流程图。

[0081] 如图6所示，该方法可以包括操作S6411～操作S6412。

[0082] 在操作S6411，使用语音识别模型获取第一字符串中被替换字符的第一语音向量和候选字符串中替换字符的第二语音向量。

[0083] 例如，语音识别模型可以是预先训练好的用于进行语音识别的神经网络模型，该模型在训练过程中可以对输入的语音信息进行特征提取，得到输入语音信息中各个字符的语音向量。经训练的模型存储有各个字符的语音向量，因此，可以直接从该语音识别模型中获取替换前后的两个字符的语音向量。

[0084] 在操作S6412，基于第一语音向量和第二语音向量之间的距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0085] 例如，第一语音向量和第二语音向量之间的距离可以是该两个语音向量之间的余弦距离，该余弦距离可以表示第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0086] 根据本公开的实施例，基于第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度来对候选字符串进行打分，考虑到了在语音识别过程中发音相似的字符被识别错误的情况，能够更精确地对候选字符串进行打分，提高候选字符串的匹配准确

率。

[0087] 图7是根据本公开的另一个实施例的信息匹配方法的流程图。

[0088] 如图7所示，该信息匹配方法700包括操作S701～S707。下面将以字符串为车牌号的情况为例对方法700进行详细说明。

[0089] 在操作S701，从输入信息中获取第一车牌号。

[0090] 例如，从用户输入的语音信息中识别出第一车牌号。

[0091] 在操作S702，将第一车牌号转换为字符序列。

[0092] 例如，将第一车牌号基于Uni-Gram切分规则切分为多个第一字符片段，并基于Bi-Gram切分规则切分为多个第二字符片段，对多个第一字符片段和多个第二字符片段中预设位置(如车牌号首部2位或尾部2未)的字符添加位置信息，基于添加位置信息后的多个第一字符片段和多个第二字符片段生成第一车牌号的字符序列。

[0093] 在操作S703，使用第一车牌号的字符序列对车牌库进行查询，得到至少两个候选车牌号。

[0094] 例如，车牌库中存储有多个第二车牌号以及与多个第二车牌号对应的多个索引序列，多个索引序列分别是对多个第二车牌号进行上述转换得到的。至少两个候选车牌号是基于字符序列分别与多个索引序列之间的相似度，从多个第二车牌号中选取的。

[0095] 在操作S704，针对每个候选车牌号计算编辑路径。

[0096] 每个候选车牌号的编辑路径表示将第一车牌号转换为该候选车牌号所需要的编辑操作的集合。具体可以计算将第一车牌号转换为该候选车牌号的最短编辑距离，确定在满足最短编辑距离的情况下将第一车牌号转换为该候选车牌号所需要的编辑操作的集合。

[0097] 在操作S705，确定增加和删除操作的扣分分数。

[0098] 在操作S706，确定替换操作的扣分分数。

[0099] 操作S705和操作S706用于确定编辑路径的扣分分数，编辑路径中所包含的编辑操作可以是增加、删除和替换，增加和删除的扣分分数可以是预先设定的数值，替换操作的扣分分数可以基于替换前后的两个字符之间的发音相似度来确定。根据编辑操作中所包含的增加、删除和替换操作的扣分分数得到编辑路径的扣分分数。

[0100] 在操作S707，计算编辑路径的得分。

[0101] 例如，可以基于预先设定的基础分数和编辑路径的扣分分数得到编辑路径的得分。

[0102] 在操作S708，计算编辑路径的得分。

[0103] 例如，可以将各个候选车牌号的编辑路径的得分作为该候选车牌号与第一车牌号之间的匹配度。

[0104] 在操作S709，确定目标车牌号。

[0105] 例如，确定与第一车牌号之间的匹配度大于0.8的候选车牌号为匹配成功的目标车牌号。

[0106] 根据本公开的实施例，相比于直接根据第一车牌号与车牌库中的第二车牌号之间的相似度来确定与第一车牌号相匹配的目标车牌号，本公开实施例基于相似度从车牌库中确定至少两个候选车牌号，基于编辑路径对至少两个候选车牌号进一步进行精细排序，从至少两个候选车牌号中确定出与第一车牌号相匹配的目标车牌号，能够提车牌号的匹配准

确率。

[0107] 图8是根据本公开的一个实施例的信息匹配装置的框图。

[0108] 如图8所示，该信息匹配装置800可以包括第一确定模块801、第二确定模块802、第三确定模块803和第四确定模块804。

[0109] 第一确定模块801用于确定第一字符串分别与预先存储的多个第二字符串之间的相似度。

[0110] 第二确定模块802用于根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串。

[0111] 第三确定模块803用于针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径。

[0112] 第四确定模块804用于根据所述编辑路径从所述候选字符串中确定至少一个候选字符串作为与所述第一字符串匹配的目标字符串。

[0113] 根据本公开的实施例，相比于直接根据第一字符串与字符串库中的第二字符串之间的相似度来确定与第一字符串相匹配的目标字符串，本公开实施例基于相似度从字符串库中确定至少两个候选字符串，基于编辑路径对至少两个候选字符串进一步进行精细排序，从至少两个候选字符串中确定出与第一字符串相匹配的目标字符串，能够提字符串的匹配准确率。

[0114] 根据本公开的实施例，第一确定模块801包括转换单元和第一确定单元。

[0115] 转换单元用于将第一字符串转换为字符序列。

[0116] 第一确定单元用于确定所述字符序列分别与多个索引序列之间的相似度作为所述第一字符串分别与多个第二字符串之间的相似度，其中所述多个索引序列分别是由所述多个第二字符串转换得到的。

[0117] 根据本公开的实施例，转换单元包括第一切分子单元、第二切分子单元、第一添加子单元、第二添加子单元和字符序列生成子单元。第一切分子单元用于将第一字符串按照第一规则切分为多个第一字符片段。第二切分子单元用于将第一字符串按照第二规则切分为多个第二字符片段。第一添加子单元用于在所述多个第一字符片段中的至少一个字符片段中，添加该第一字符片段在第一字符串中的位置信息，得到第一字符片段的序列。第二添加子单元用于在所述多个第二字符片段中的至少二个字符片段中，添加该第二字符片段在第一字符串中的位置信息，得到第二字符片段的序列。字符序列生成子单元用于基于第一字符片段的序列和第二字符片段的序列，生成所述字符序列。

[0118] 根据本公开的实施例，所述第一规则包括以单个字符作为一个切分单元进行切分，所述第二规则包括以任意两个相邻字符作为切分单元进行切分。例如第一规则可以包括Uni-Gram切分规则，所述第二规则可以包括Bi-Gram切分规则。

[0119] 根据本公开的实施例，第二确定模块802是在多个索引序列中不存在与字符序列相同的索引序列的情况下执行所述根据所述相似度从所述多个第二字符串中确定至少两个第二字符串作为候选字符串的操作。

[0120] 根据本公开的实施例，第三确定模块803包括第一计算单元和第二确定单元。

[0121] 第一计算单元用于计算第一字符串与候选字符串之间的最短编辑距离，所述最短编辑距离表示将第一字符串转换为候选字符串的最少操作次数。

[0122] 第二确定单元用于确定在满足最短编辑距离的情况下将第一字符串转换为候选字符串的编辑操作集合，作为编辑路径。

[0123] 根据本公开的实施例，第四确定模块804包括打分单元、第二计算单元和第三确定单元。

[0124] 打分单元用于根据每个候选字符串的编辑路径中包含的编辑操作的类型和数量，对所述编辑路径进行打分。

[0125] 第二计算单元用于根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度。

[0126] 第三确定单元用于从所述候选字符串中确定具有符合预设条件的匹配度的至少一个候选字符串，作为与所述第一字符串匹配的目标字符串。

[0127] 根据本公开的实施例，第二计算单元具体用于在候选字符串存在多个编辑路径的情况下，根据多个编辑路径的得分中的最高分来计算该候选字符串与第一字符串之间的匹配度。

[0128] 根据本公开的实施例，打分单元包括第一确定子单元和第二确定子单元。

[0129] 第一确定子单元用于根据所述编辑路径中包含的编辑操作的类型和数量来确定所述编辑路径的扣分分数。

[0130] 第二确定子单元用于根据针对所述编辑路径而预先设置的基础分数和所述编辑路径的扣分分数，确定所述编辑路径的得分。

[0131] 根据本公开的实施例，所述第一字符串是从语音输入信息中识别得到的。第一确定子单元包括扣分确定子单元和扣分计算子单元。扣分确定子单元用于针对编辑路径中包含的每个编辑操作，在所述编辑操作为删除操作的情况下，将所述删除操作的扣分分数确定为第一预设值；在所述编辑操作为增加操作的情况下，将所述增加操作的扣分分数确定为第二预设值；在所述编辑操作为替换操作的情况下，根据第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度，确定所述替换操作的扣分分数。扣分计算子单元用于基于编辑路径中包含的至少一个编辑操作的扣分分数，来计算所述编辑路径的扣分分数。

[0132] 根据本公开的实施例，扣分确定子单元包括发音相似度计算子单元和替换操作扣分子单元。发音相似度计算子单元用于计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。替换操作扣分子单元根据以下等式计算所述编辑路径中的替换操作的扣分分数：$q=-A*(1-p)$其中，q表示所述编辑路径中的替换操作的扣分分数，A为预设参数，p表示第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0133] 根据本公开的实施例，发音相似度计算子单元包括拼音映射子单元、距离计算子单元和第一发音相似度确定子单元。拼音映射子单元用于基于预设的拼音编码映射表，生成第一字符串中被替换字符的第一拼音编码以及候选字符串中替换字符的第二拼音编码。距离计算子单元用于计算第一拼音编码和第二拼音编码之间的最短编辑距离。第一发音相似度确定子单元用于根据第一拼音编码和第二拼音编码之间的最短编辑距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0134] 根据本公开的实施例，发音相似度计算子单元包括语音识别子单元和第二发音相

似度确定子单元。语音识别子单元用于使用语音识别模型获取第一字符串中被替换字符的第一语音向量和候选字符串中替换字符的第二语音向量。第二发音相似度确定子单元用于基于第一语音向量和第二语音向量之间的距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度。

[0135] 根据本公开的实施例，本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。根据本公开的实施例，相比于直接根据第一字符串与字符串库中的第二字符串之间的相似度来确定与第一字符串相匹配的目标字符串，本公开实施例基于相似度从字符串库中确定至少两个候选字符串，基于编辑路径对至少两个候选字符串进一步进行精细排序，从至少两个候选字符串中确定出与第一字符串相匹配的目标字符串，能够提字符串的匹配准确率。

[0136] 图9示出了可以用来实施本公开的实施例的示例电子设备900的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0137] 如图9所示，设备900包括计算单元901，其可以根据存储在只读存储器(ROM)902中的计算机程序或者从存储单元908加载到随机访问存储器(RAM)903中的计算机程序，来执行各种适当的动作和处理。在RAM 903中，还可存储设备900操作所需的各种程序和数据。计算单元901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O)接口905也连接至总线904。

[0138] 设备900中的多个部件连接至I/O接口905，包括：输入单元906，例如键盘、鼠标等；输出单元907，例如各种类型的显示器、扬声器等；存储单元908，例如磁盘、光盘等；以及通信单元909，例如网卡、调制解调器、无线通信收发机等。通信单元909允许设备900通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0139] 计算单元901可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元901的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元901执行上文所描述的各个方法和处理，例如信息匹配方法。例如，在一些实施例中，信息匹配方法可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元908。在一些实施例中，计算机程序的部分或者全部可以经由ROM 902和/或通信单元909而被载入和/或安装到设备900上。当计算机程序加载到RAM 903并由计算单元901执行时，可以执行上文描述的信息匹配方法的一个或多个步骤。备选地，在其他实施例中，计算单元901可以通过其他任何适当的方式(例如，借助于固件)而被配置为执行信息匹配方法。

[0140] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算

机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0141] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0142] 在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0143] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置(例如，CRT(阴极射线管)或者LCD(液晶显示器)监视器)；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0144] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)和互联网。

[0145] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。

[0146] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本公开公开的技术方案所期望的结果，本文在此不进行限制。

[0147] 上述具体实施方式，并不构成对本公开保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等，均应包含在本公开保护范围之内。

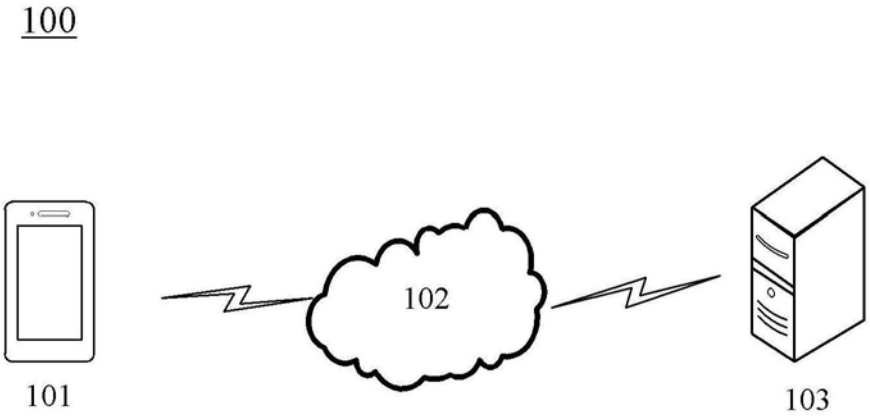

100
101
102
103

图1

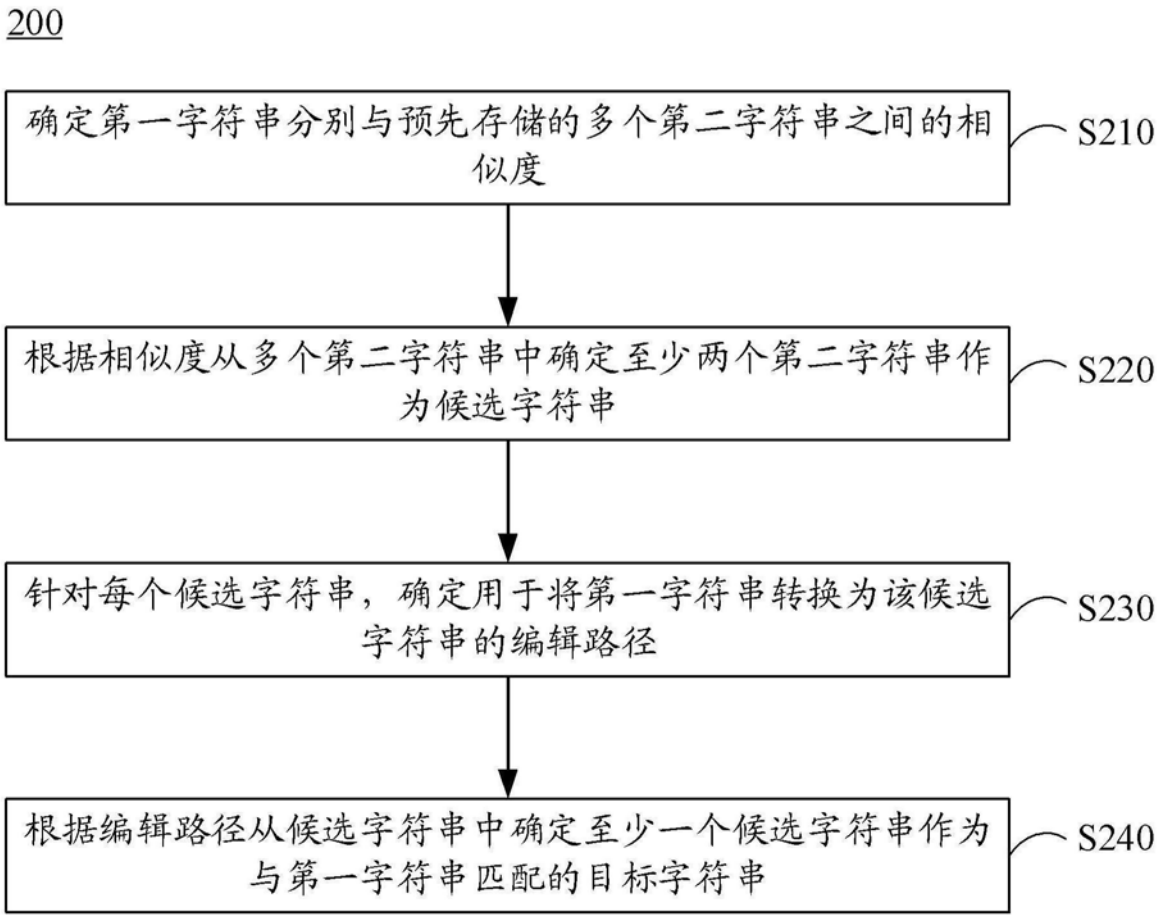

200
确定第一字符串分别与预先存储的多个第二字符串之间的相似度
S210
根据相似度从多个第二字符串中确定至少两个第二字符串作为候选字符串
S220
针对每个候选字符串，确定用于将第一字符串转换为该候选字符串的编辑路径
S230
根据编辑路径从候选字符串中确定至少一个候选字符串作为与第一字符串匹配的目标字符串
S240

图2

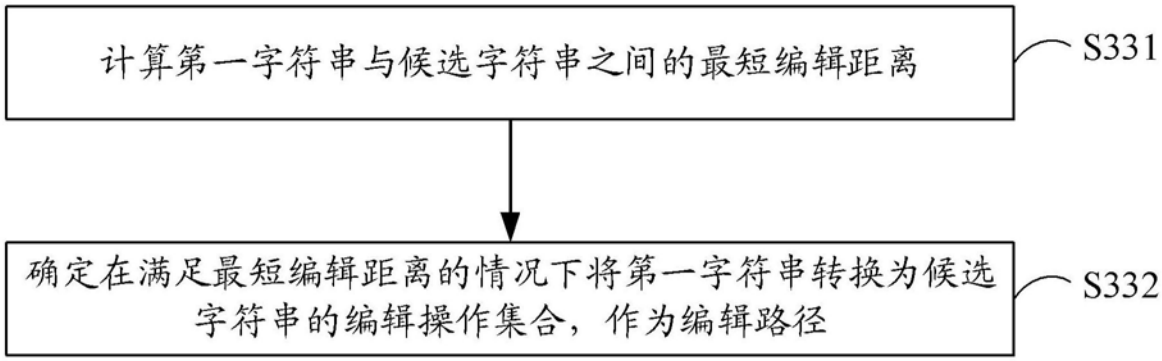

计算第一字符串与候选字符串之间的最短编辑距离
S331
确定在满足最短编辑距离的情况下将第一字符串转换为候选字符串的编辑操作集合，作为编辑路径
S332

图3

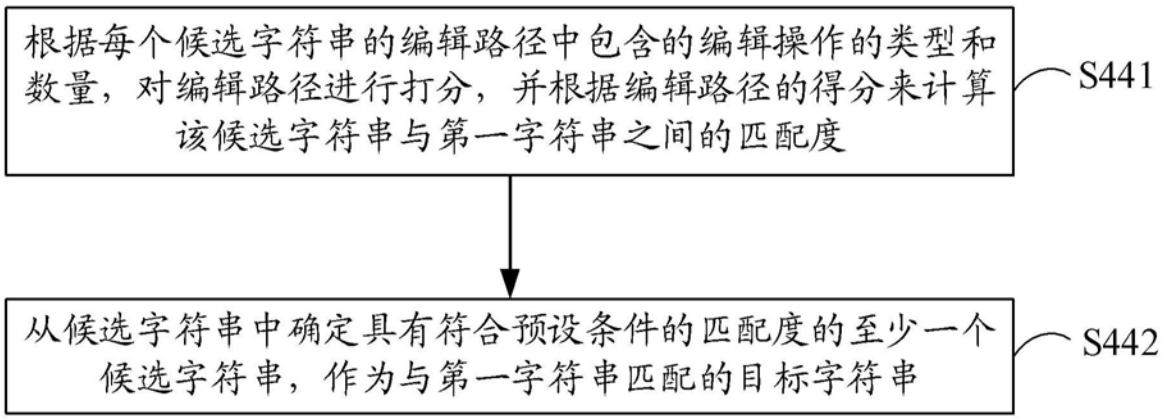

根据每个候选字符串的编辑路径中包含的编辑操作的类型和数量，对编辑路径进行打分，并根据编辑路径的得分来计算该候选字符串与第一字符串之间的匹配度
S441
从候选字符串中确定具有符合预设条件的匹配度的至少一个候选字符串，作为与第一字符串匹配的目标字符串
S442

图4

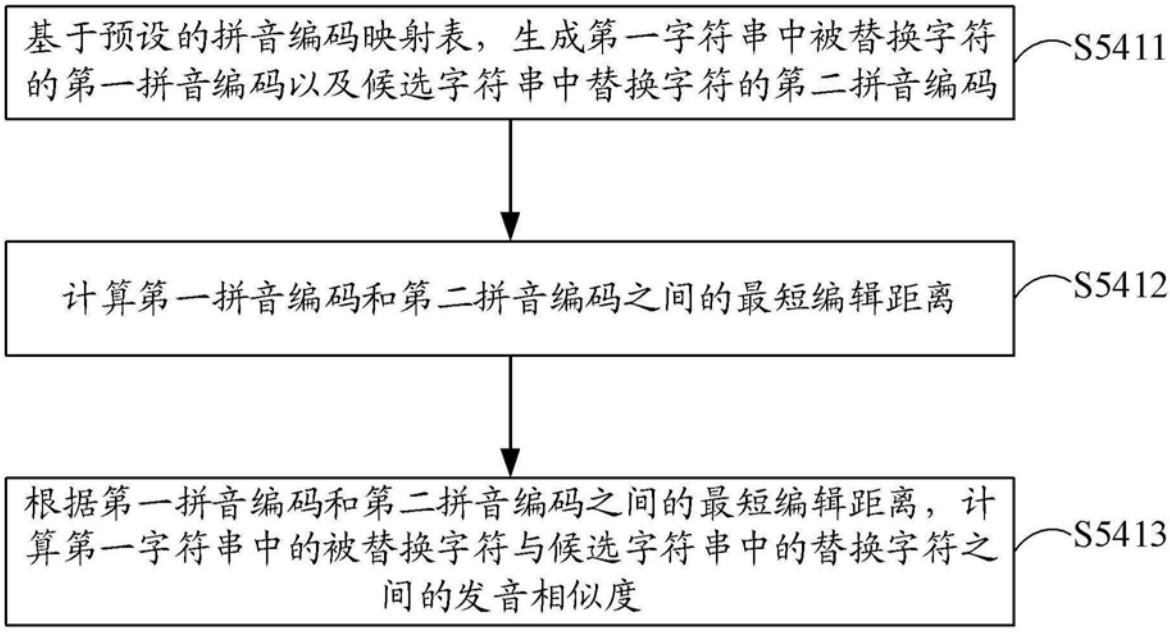

基于预设的拼音编码映射表，生成第一字符串中被替换字符的第一拼音编码以及候选字符串中替换字符的第二拼音编码
S5411
计算第一拼音编码和第二拼音编码之间的最短编辑距离
S5412
根据第一拼音编码和第二拼音编码之间的最短编辑距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度
S5413

图5

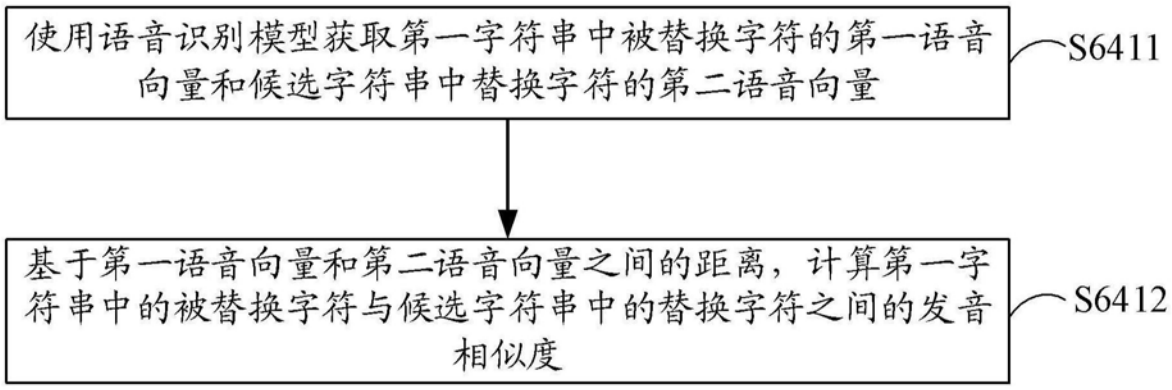

使用语音识别模型获取第一字符串中被替换字符的第一语音向量和候选字符串中替换字符的第二语音向量
S6411
基于第一语音向量和第二语音向量之间的距离，计算第一字符串中的被替换字符与候选字符串中的替换字符之间的发音相似度
S6412

图6

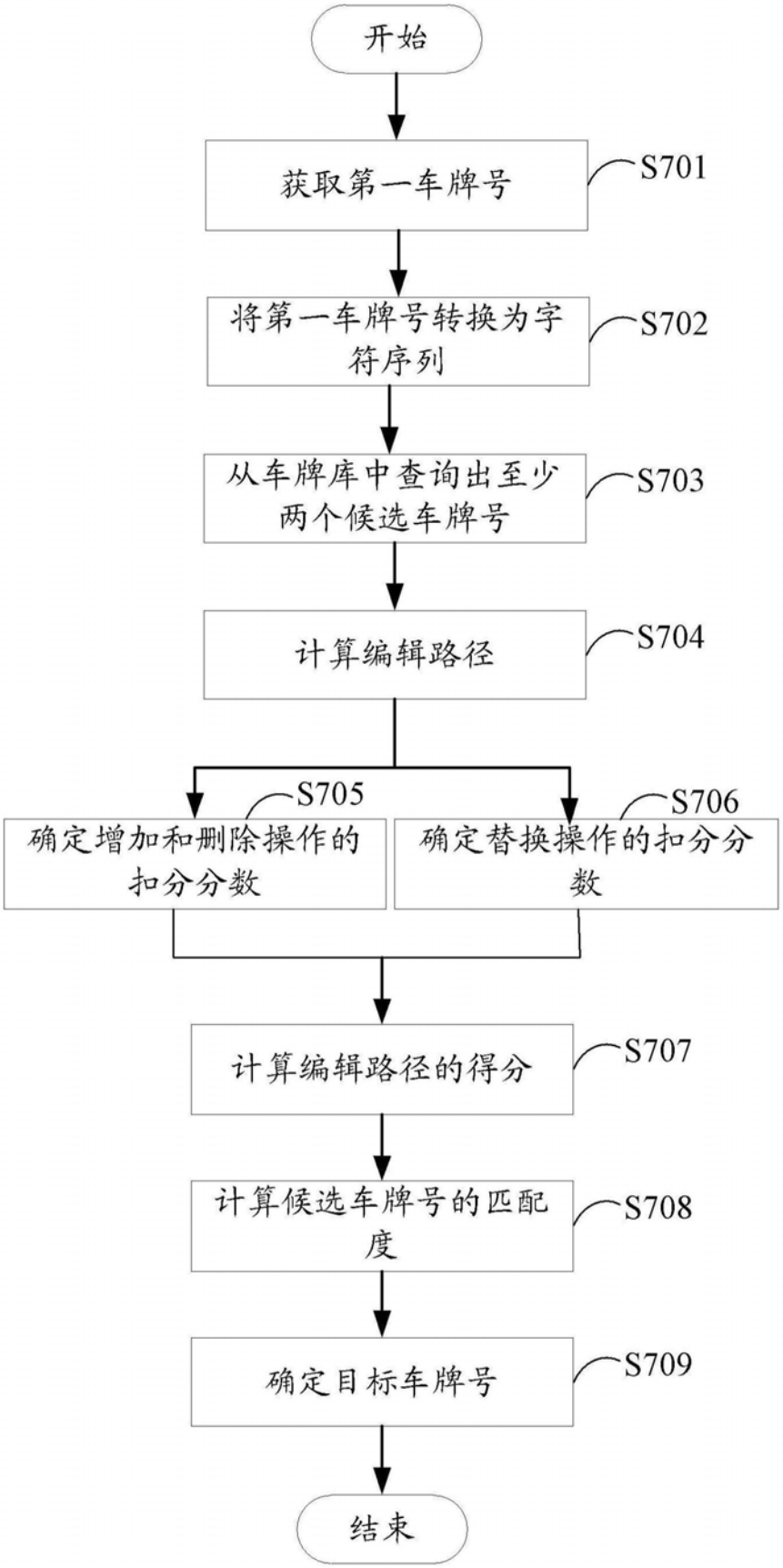

开始
获取第一车牌号
S701
将第一车牌号转换为字符序列
S702
从车牌库中查询出至少两个候选车牌号
S703
计算编辑路径
S704
S705
确定增加和删除操作的扣分分数
S706
确定替换操作的扣分分数
计算编辑路径的得分
S707
计算候选车牌号的匹配度
S708
确定目标车牌号
S709
结束

图7

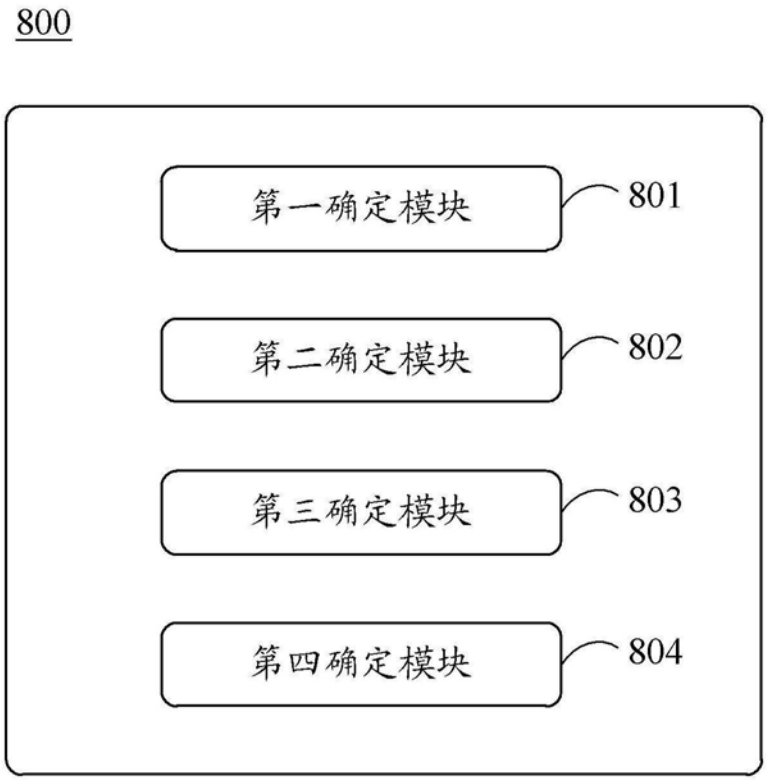

800
第一确定模块
801
第二确定模块
802
第三确定模块
803
第四确定模块
804

图8

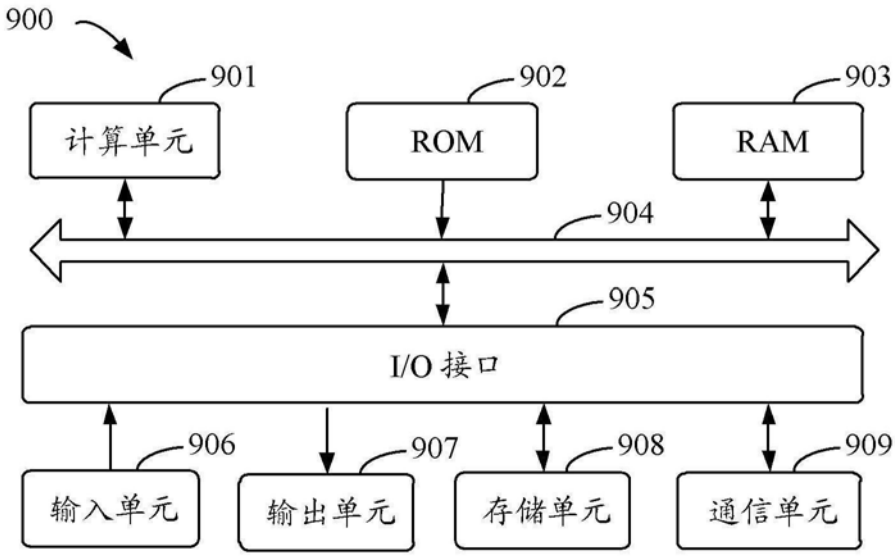

900
901
计算单元
902
ROM
903
RAM
904
905
I/O 接口
906
输入单元
907
输出单元
908
存储单元
909
通信单元

图9