(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 111862946 A
(43)申请公布日 2020.10.30

(21)申请号 201910414644.8

(22)申请日 2019.05.17

(71)申请人 北京嘀嘀无限科技发展有限公司
地址 100193 北京市海淀区东北旺西路8号院34号楼

(72)发明人 葛檬 张睿雄

(74)专利代理机构 北京超成律师事务所 11646
代理人 刘静

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/26(2006.01)
G10L 25/03(2013.01)
G10L 25/48(2013.01)

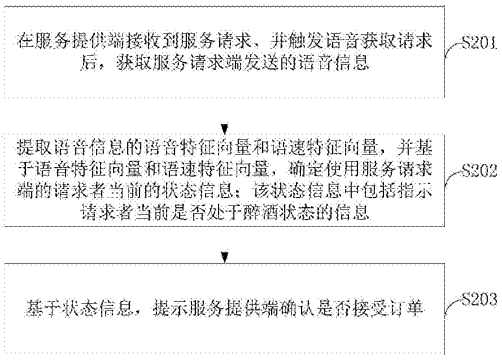
权利要求书3页 说明书20页 附图5页

(54)发明名称

一种订单处理方法、装置、电子设备及存储介质

(57)摘要

本申请提供了一种订单处理方法、装置、电子设备及存储介质，其中，该订单处理方法包括：在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息；提取所述语音信息的语音特征向量和语速特征向量，并基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息；所述状态信息中包括指示所述请求者当前是否处于醉酒状态的信息；基于所述状态信息，提示所述服务提供端确认是否接受所述订单。本申请能够提高乘车环境的安全性。



CN 111862946 A

1.一种订单处理装置,其特征在于,包括:

获取模块,用于在服务提供端接收到服务请求、并触发语音获取请求后,获取服务请求端发送的语音信息,并将所述语音信息传输至确定模块;

确定模块,用于提取所述语音信息的语音特征向量和语速特征向量,并基于所述语音特征向量和所述语速特征向量,确定使用所述服务请求端的请求者当前的状态信息,并将所述状态信息传输至提示模块;所述状态信息中包括指示所述请求者当前是否处于醉酒状态的信息;

提示模块,用于基于所述状态信息,提示所述服务提供端确认是否接受所述订单。

2.根据权利要求1所述的订单处理装置,其特征在于,还包括处理模块,所述处理模块在所述获取模块获取服务请求端发送的语音信息之后,在所述确定模块提取所述语音信息的语音特征向量和语速特征向量之前,用于:

对所述语音信息进行语音端点检测,得到至少一个语音段落和静音段落;

删除所述语音信息中的静音段落。

3.根据权利要求2所述的订单处理装置,其特征在于,所述确定模块具体用于根据以下步骤提取所述语音信息时的特征向量:

对所述语音信息中的各个语音段落分别进行分帧处理,得到每个语音段落对应的语音帧;

针对每个语音段落,提取该语音段落中每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差,并基于所述语音帧特征、所述语音帧特征差和预先设定的语音段落特征函数,确定该语音段落的第一段落语音特征向量;

基于所述语音信息的各个语音段落分别对应的第一段落语音特征向量,提取所述语音信息的语音特征向量。

4.根据权利要求3所述的订单处理装置,其特征在于,所述确定模块具体用于根据以下步骤基于所述语音信息的各个语音段落分别对应的第一段落语音特征向量,提取所述语音信息的语音特征向量:

针对每个语音段落,基于该语音段落的所述第一段落语音特征向量和预先存储的清醒状态语音特征向量,确定每个语音段落的差异性语音特征向量;

基于每个语音段落的所述第一段落语音特征向量和所述差异性语音特征向量,确定每个语音段落的第二段落语音特征向量;

对各个第二段落语音特征向量进行合并,得到所述语音信息的语音特征向量。

5.根据权利要求2所述的订单处理装置,其特征在于,所述确定模块具体用于根据以下步骤提取所述语音信息的语速特征向量:

将所述语音信息中的各个语音段落转换为文本段落,每个文本段落包括多个字符;

基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长,确定每个文本段落的语速;

基于每个文本段落对应的语速,确定所述语音信息的最大语速、最小语速和平均语速;

基于所述语音信息的最大语速、最小语速和平均语速,提取所述语音信息的语速特征向量。

6.根据权利要求2所述的订单处理装置,其特征在于,所述确定模块具体用于根据以下

步骤基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息：

基于所述语音特征向量确定所述语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量，所述第一分值特征向量包括所述语音信息中每个语音段落指示醉酒状态的概率值，所述第二分值特征向量包括所述语音信息中每个语音段落指示不醉酒状态的概率值；

基于所述第一分值特征向量、所述第二分值特征向量以及所述语速特征向量，确定所述请求者当前的状态信息。

7.根据权利要求6所述的订单处理装置，其特征在于，所述确定模块具体用于根据以下步骤基于所述语音特征向量确定所述语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量：

将所述语音特征向量输入预先训练的声音识别模型中的段级别分类器，得到所述语音信息指示所述醉酒状态的第一分值特征向量和指示所述不醉酒状态的第二分值特征向量；

基于所述第一分值特征向量、所述第二分值特征向量和预先设定的语音分值特征函数，确定所述语音信息的分值特征向量；

将所述分值特征向量和所述语速特征向量合并后，输入所述声音识别模型中的状态级别分类器，确定所述请求者当前的状态信息。

8.根据权利要求7所述的订单处理装置，其特征在于，还包括模型训练模块，所述模型训练模块用于根据以下步骤训练所述声音识别模型：

构建声音识别模型的段级别分类器和状态级别分类器；

获取预先构建的训练样本库，所述训练样本库包括多个训练语音信息对应的训练语音特征向量、训练语速特征向量以及每个训练语音信息对应的状态信息；

依次将每个训练语音信息的训练语音特征向量输入所述段级别分类器，得到该训练语音信息对应的第一训练分值特征向量和第二训练分值特征向量；将所述第一训练分值特征向量、所述第二训练分值特征向量和所述训练语速特征向量作为所述状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为所述声音识别模型的输出变量，训练得到所述声音识别模型的模型参数信息。

9.根据权利要求8所述的订单处理装置，其特征在于，还包括模型测试模块，所述模型测试模块用于根据以下步骤测试所述声音识别模型：

获取预先构建的测试样本库，所述测试样本库包括多个测试语音信息对应的测试语音特征向量、测试语速特征向量以及每个测试语音信息对应的真实状态信息；

依次将所述测试样本库中的每个测试语音特征向量输入所述声音识别模型的段级分类器，得到所述测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量；

将所述测试样本库中每个测试语音对应的第一测试分值特征向量、第二测试分值特征向量和测试语速特征向量输入所述声音识别模型的状态级别分类器，得到所述测试样本库中的每个测试语音对应的测试状态信息；

基于所述真实状态信息和所述测试状态信息，确定所述声音识别模型的精确率和召回率。

10.一种订单处理方法，其特征在于，包括：

在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息；

提取所述语音信息的语音特征向量和语速特征向量，并基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息；所述状态信息中包括指示所述请求者当前是否处于醉酒状态的信息；

基于所述状态信息，提示所述服务提供端确认是否接受所述订单。

11.一种电子设备，其特征在于，包括：处理器、存储介质和总线，所述存储介质存储有所述处理器可执行的机器可读指令，当电子设备运行时，所述处理器与所述存储介质之间通过总线通信，所述处理器执行所述机器可读指令，以执行如权利要求10所述订单处理方法的步骤。

12.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器运行时执行如权利要求10所述订单处理方法的步骤。

一种订单处理方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及计算机技术领域，具体而言，涉及一种订单处理方法、装置、电子设备及存储介质。

背景技术

[0002] 随着互联网移动通信技术及智能设备的迅速发展，各种服务类应用程序也随之诞生，比如用车应用(Application，APP)。服务请求可以通过用车APP请求获取相应的用车服务，当用车平台接收到服务请求方发起的出行请求时，会为用户匹配服务提供方，提供相应的出行服务。

[0003] 在订单派送中，服务提供方具有选择接受订单或者拒绝订单的权利。研究中发现，服务请求方在醉酒状态下独自请求出行服务时出现事故的概率比较大，比如服务请求方醉酒闹事影响服务提供方的正常驾驶，或者，出现威胁到服务提供方人身安全的情况。目前服务提供方只能在服务请求方上车后通过人为判断是否存在醉酒情况，无法在服务请求方上车前进行预判，从而无法对乘车环境的安全性提前进行风险控制。

发明内容

[0004] 有鉴于此，本申请的目的在于提供一种订单处理方法、装置、电子设备及存储介质，能够对乘车环境的安全性提前进行风险控制，进而提高了整体乘车环境的安全性。

[0005] 第一方面，本申请实施例提供了一种订单处理装置，包括：

[0006] 获取模块，用于在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息，并将所述语音信息传输至确定模块；

[0007] 确定模块，用于提取所述语音信息的语音特征向量和语速特征向量，并基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息，并将所述状态信息传输至提示模块；所述状态信息中包括指示所述请求者当前是否处于醉酒状态的信息；

[0008] 提示模块，用于基于所述状态信息，提示所述服务提供端确认是否接受所述订单。

[0009] 在一些实施方式中，还包括处理模块，所述处理模块在所述获取模块获取服务请求端发送的语音信息之后，在所述确定模块提取所述语音信息的语音特征向量和语速特征向量之前，用于：

[0010] 对所述语音信息进行语音端点检测，得到至少一个语音段落和静音段落；

[0011] 删除所述语音信息中的静音段落。

[0012] 在一些实施方式中，所述确定模块具体用于根据以下步骤提取所述语音信息时的特征向量：

[0013] 对所述语音信息中的各个语音段落分别进行分帧处理，得到每个语音段落对应的语音帧；

[0014] 针对每个语音段落，提取该语音段落中每个语音帧的语音帧特征以及该语音帧与

其相邻的语音帧之间的语音帧特征差，并基于所述语音帧特征、所述语音帧特征差和预先设定的语音段落特征函数，确定该语音段落的第一段落语音特征向量；

[0015] 基于所述语音信息的各个语音段落分别对应的第一段落语音特征向量，提取所述语音信息的语音特征向量。

[0016] 在一些实施方式中，所述确定模块具体用于根据以下步骤基于所述语音信息的各个语音段落分别对应的第一段落语音特征向量，提取所述语音信息的语音特征：

[0017] 针对每个语音段落，基于该语音段落的所述第一段落语音特征向量和预先存储的清醒状态语音特征向量，确定每个语音段落的差异性语音特征向量；

[0018] 基于每个语音段落的所述第一段落语音特征向量和所述差异性语音特征向量，确定每个语音段落的第二段落语音特征向量；

[0019] 对各个第二段落语音特征向量进行合并，得到所述语音信息的语音特征向量。

[0020] 在一些实施方式中，所述确定模块具体用于根据以下步骤提取所述语音信息的语速特征向量：

[0021] 将所述语音信息中的各个语音段落转换为文本段落，每个文本段落包括多个字符；

[0022] 基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长，确定每个文本段落的语速；

[0023] 基于每个文本段落对应的语速，确定所述语音信息的最大语速、最小语速和平均语速；

[0024] 基于所述语音信息的最大语速、最小语速和平均语速，提取所述语音信息的语速特征向量。

[0025] 在一些实施方式中，所述确定模块具体用于根据以下步骤基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息：

[0026] 基于所述语音特征向量确定所述语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量，所述第一分值特征向量包括所述语音信息中每个语音段落指示醉酒状态的概率值，所述第二分值特征向量包括所述语音信息中每个语音段落指示不醉酒状态的概率值；

[0027] 基于所述第一分值特征向量、所述第二分值特征向量以及所述语速特征向量，确定所述请求者当前的状态信息。

[0028] 在一些实施方式中，所述确定模块具体用于根据以下步骤基于所述语音特征向量确定所述语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量：

[0029] 将所述语音特征向量输入预先训练的声音识别模型中的段级别分类器，得到所述语音信息指示所述醉酒状态的第一分值特征向量和指示所述不醉酒状态的第二分值特征向量；

[0030] 基于所述第一分值特征向量、所述第二分值特征向量和预先设定的语音分值特征函数，确定所述语音信息的分值特征向量；

[0031] 将所述分值特征向量和所述语速特征向量合并后，输入所述声音识别模型中的状态级别分类器，确定所述请求者当前的状态信息。

[0032] 在一些实施方式中，还包括模型训练模块，所述模型训练模块用于根据以下步骤训练所述声音识别模型：

[0033] 构建声音识别模型的段级别分类器和状态级别分类器；

[0034] 获取预先构建的训练样本库，所述训练样本库包括多个训练语音信息对应的训练语音特征向量、训练语速特征向量以及每个训练语音信息对应的状态信息；

[0035] 依次将每个训练语音信息的训练语音特征向量输入所述段级别分类器，得到该训练语音信息对应的第一训练分值特征向量和第二训练分值特征向量；将所述第一训练分值特征向量、所述第二训练分值特征向量和所述训练语速特征向量作为所述状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为所述声音识别模型的输出变量，训练得到所述声音识别模型的模型参数信息。

[0036] 在一些实施方式中，还包括模型测试模块，所述模型测试模块用于根据以下步骤测试所述声音识别模型：

[0037] 获取预先构建的测试样本库，所述测试样本库包括多个测试语音信息对应的测试语音特征向量、测试语速特征向量以及每个测试语音信息对应的真实状态信息；

[0038] 依次将所述测试样本库中的每个测试语音特征向量输入所述声音识别模型的段级分类器，得到所述测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量；

[0039] 将所述测试样本库中每个测试语音对应的第一测试分值特征向量、第二测试分值特征向量和测试语速特征向量输入所述声音识别模型的状态级别分类器，得到所述测试样本库中的每个测试语音对应的测试状态信息；

[0040] 基于所述真实状态信息和所述测试状态信息，确定所述声音识别模型的精确率和召回率；

[0041] 若所述精确率和所述召回率不满足设定条件，更新所述声音识别模型中的模型训练参数和/或所述训练样本库，重新训练所述声音识别模型，直至所述精确率和所述召回率满足所述设定条件。

[0042] 第二方面，本申请实施例提供了一种订单处理方法，包括：

[0043] 在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息；

[0044] 提取所述语音信息的语音特征向量和语速特征向量，并基于所述语音特征向量和所述语速特征向量，确定使用所述服务请求端的请求者当前的状态信息；所述状态信息中包括指示所述请求者当前是否处于醉酒状态的信息；

[0045] 基于所述状态信息，提示所述服务提供端确认是否接受所述订单。

[0046] 第三方面，本申请实施例提供了一种电子设备，包括：处理器、存储介质和总线，所述存储介质存储有所述处理器可执行的机器可读指令，当电子设备运行时，所述处理器与所述存储介质之间通过总线通信，所述处理器执行所述机器可读指令，以执行如第二方面所述订单处理方法的步骤。

[0047] 第四方面，本申请实施例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上存储有计算机程序，所述计算机程序被处理器运行时执行如第一方面所述订单处理方法的步骤。

[0048] 本申请实施例提供了一种订单处理方法、装置、服务器及计算机可读存储介质，通过在服务提供端接收到服务请求、并触发语音获取请求后，获取到服务请求端发送的语音信息，然后提取该语音信息的语音特征向量和语速特征向量，并基于该语音特征向量和语速特征向量确定服务请求端的请求者当前的状态信息，即确定请求者是否处于醉酒状态的信息，然后基于请求者是否处于醉酒状态的信息，提示服务提供端确认是否接受订单，这样通过提前确定请求者是否处于醉酒状态，并对服务提供端进行相应的提示，从而对乘车环境的安全性提前进行风险控制，进而提高了整体乘车环境的安全性。

[0049] 本申请实施例的其他特征和优点将在随后的说明书中阐述，或者，部分特征和优点可以从说明书推知或毫无疑义地确定，或者通过实施本申请实施例的上述技术即可得知。

[0050] 为使本申请的上述目的、特征和优点能更明显易懂，下文特举较佳实施例，并配合所附附图，作详细说明如下。

附图说明

[0051] 为了更清楚地说明本申请实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本申请的某些实施例，因此不应被看作是对范围的限定，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他相关的附图。

[0052] 图1示出了本申请实施例提供的一种订单处理系统的架构示意图；

[0053] 图2示出了本申请实施例提供的一种订单处理方法的流程图；

[0054] 图3示出了本申请实施例提供的第一种提取语音信息的语音特征向量的方法流程图；

[0055] 图4示出了本申请实施例提供的第二种提取语音信息的语音特征向量的方法流程图；

[0056] 图5示出了本申请实施例提供的一种提取语音信息的语速特征向量的方法流程图；

[0057] 图6示出了本申请实施例提供的一种基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息的方法流程图；

[0058] 图7示出了本申请实施例提供的一种声音识别模型的精确率-召回率曲线示意图；

[0059] 图8示出了本申请实施例提供的一种订单处理装置的结构示意图；

[0060] 图9示出了本申请实施例提供的一种电子设备的结构示意图。

具体实施方式

[0061] 为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，应当理解，本申请中附图仅起到说明和描述的目的，并不用于限定本申请的保护范围。另外，应当理解，示意性的附图并未按实物比例绘制。本申请中使用的流程图示出了根据本申请的一些实施例实现的操作。应该理解，流程图的操作可以不按顺序实现，没有逻辑的上下文关系的步骤可以反转顺序或者同时实施。此外，本领域技术人员在本申请内容的指引下，可以向流程图添加一个或

多个其他操作，也可以从流程图中移除一个或多个操作。

[0062] 另外，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围，而是仅仅表示本申请的选定实施例。基于本申请的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0063] 为了使得本领域技术人员能够使用本申请内容，结合特定应用场景“网约车订单处理”，给出以下实施方式。对于本领域技术人员来说，在不脱离本申请的精神和范围的情况下，可以将这里定义的一般原理应用于其他实施例和应用场景。虽然本申请主要围绕出网约车订单进行描述，但是应该理解，这仅是一个示例性实施例。

[0064] 需要说明的是，本申请实施例中将会用到术语“包括”，用于指出其后所声明的特征的存在，但并不排除增加其它的特征。

[0065] 本申请中的术语“乘客”、“请求方”、“请求者”、“服务请求方”和“客户”可互换使用，以指代可以请求或订购服务的个人、实体或工具。本申请中的术语“司机”、“提供方”、“提供者”、“服务提供方”和“供应商”可互换使用，以指代可以提供服务的个人、实体或工具。本申请中的术语“用户”可以指代请求服务、订购服务、提供服务或促成服务的提供的个人、实体或工具。例如，用户可以是乘客、驾驶员、操作员等，或其任意组合。在本申请中，“乘客”和“乘客终端”可以互换使用，“驾驶员”和“驾驶员终端”可以互换使用。

[0066] 本申请中的术语“服务请求”和“订单”可互换使用，以指代由乘客、服务请求方、司机、服务提供方、或供应商等、或其任意组合发起的请求。接受该“服务请求”或“订单”的可以是乘客、服务请求方、司机、服务提供方、或供应商等、或其任意组合。服务请求可以是收费的或免费的。

[0067] 本申请中使用的定位技术可以基于全球定位系统(Global Positioning System,GPS)、全球导航卫星系统(Global Navigation Satellite System,GLONASS)，罗盘导航系统(COMPASS)、伽利略定位系统、准天顶卫星系统(Quasi-Zenith Satellite System,QZSS)、无线保真(Wireless Fidelity,WiFi)定位技术等，或其任意组合。一个或多个上述定位系统可以在本申请中互换使用。

[0068] 本申请的一个方面涉及一种订单处理系统。该系统可以通过对服务请求端发送的语音信息进行处理，确定使用服务请求端的请求者当前是否处于醉酒状态，并根据确定的请求者的当前状态，提示服务提供端确认是否接受服务请求端发送的订单。

[0069] 值得注意的是，在本申请提出申请之前，服务提供方只能在服务请求方上车后通过人为判断是否存在醉酒情况，无法在服务请求方上车前进行预判，从而无法对乘车环境的安全性提前进行风险控制。然而，本申请提供的订单处理系统可以在服务请求方上车之前判断服务请求方是否存在醉酒情况，进而提前对服务请求端进行提示。因此，通过提前对服务请求端进行提示，本申请的订单处理系统可以为乘车环境的安全性提前进行风险控制。

[0070] 图1是本申请实施例提供的一种订单处理系统100的架构示意图。例如，订单处理系统100可以是用于诸如出租车、代驾服务、快车、拼车、公共汽车服务、驾驶员租赁、或班车服务之类的运输服务、或其任意组合的在线运输服务平台。订单处理系统100可以包括服务

器101、网络102、服务请求端103、服务提供端104、和数据库105中的一种或多种。

[0071] 在一些实施例中，服务器101可以包括处理器。处理器可以处理与服务请求有关的信息和/或数据，以执行本申请中描述的一个或多个功能。在一些实施例中，处理器可以包括一个或多个处理核(例如，单核处理器(S)或多核处理器(S))。仅作为举例，处理器可以包括中央处理单元(Central Processing Unit,CPU)、专用集成电路(Application Specific Integrated Circuit,ASIC)、专用指令集处理器(Application Specific Instruction-set Processor,ASIP)、图形处理单元(Graphics Processing Unit,GPU)、物理处理单元(Physics Processing Unit,PPU)、数字信号处理器(Digital Signal Processor,DSP)、现场可编程门阵列(Field Programmable Gate Array,FPGA)、可编程逻辑器件(Programmable Logic Device,PLD)、控制器、微控制器单元、简化指令集计算机(Reduced Instruction Set Computing,RISC)、或微处理器等，或其任意组合。

[0072] 在一些实施例中，服务请求端103和服务提供端104对应的设备类型可以是移动设备，比如可以包括智能家居设备、可穿戴设备、智能移动设备、虚拟现实设备、或增强现实设备等，也可以是平板计算机、膝上型计算机、或机动车辆中的内置设备等。

[0073] 在一些实施例中，数据库105可以连接到网络102以与订单处理系统100中的一个或多个组件(例如，服务器101，服务请求端103，服务提供端104等)通信。订单处理系统100中的一个或多个组件可以经由网络102访问存储在数据库105中的数据或指令。在一些实施例中，数据库150可以直接连接到订单处理系统100中的一个或多个组件，或者，数据库105也可以是服务器101的一部分。

[0074] 下面结合上述图1示出的订单处理系统100中描述的内容，对本申请实施例提供的订单处理方法进行详细说明。

[0075] 参照图2所示，为本申请实施例提供的一种订单处理方法的流程示意图，该方法可以由订单处理系统100中的服务器或服务提供端来执行，具体执行过程包括以下步骤S201～S203：

[0076] S201，在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息。

[0077] 在出行领域，这里的服务提供端可以是司机端的移动设备，服务请求端可以是乘客端的移动设备，这里的移动设备比如可以包括智能家居设备、可穿戴设备、智能移动设备、虚拟现实设备、或增强现实设备等，也可以是平板计算机、膝上型计算机、或机动车辆中的内置设备等。

[0078] 这里的服务请求可以理解为用车请求或者订单请求，比如乘客通过移动终端上的用车应用程序(Application,APP)发送用车请求，该用车请求可以保护乘客的当前位置信息以及通讯地址，比如移动终端的号码。

[0079] 司机端在接收到乘客端发送的服务请求时，可以通过服务请求中的通讯地址触发语音获取请求，当司机端与乘客端建立语音通话连接时，服务器或者司机端可以获取服务请求端即乘客端发送的语音信息。

[0080] 本申请实施例中，这里的语音信息主要指乘客的语音信息，即获取乘客在司机端与乘客端建立语音通话连接后，获取的乘客的语音信息，比如包括乘客叙述的自己的当前位置，以及要去的目的地等信息。

[0081] 由于乘客在叙述自己的位置以及要去的目的地时，语音信息中可能会包含一部分不含语音的静音段落，去除掉这部分静音段落后，能够提高后续基于语音信息识别乘客的状态信息，且也能防止一些脏数据的侵入，故在获取到乘客端的语音信息后，本申请实施例的订单处理方法还包括：

[0082] (1)对语音信息进行语音端点检测，得到至少一个语音段落和静音段落；

[0083] (2)删除语音信息中的静音段落。

[0084] 这里通过语音端点检测(Voice Activity Detection,VAD)能够去除语音信息中的静音段落，比如删除由于乘客正在听对方说话、思考、稍事休息等原因引起的一段话之间的停顿导致的空白段落，或者明显不属于乘客的语音信息的噪音段落，比如汽车鸣笛、风声、雨声等噪音段落。

[0085] 当语音信息内容比较长时，删除静音段落后，可以得到多个语句信息，然后可以按照设定长度对语句信息进行分段，得到多个语音段落，当语音信息内容比较短时，在删除掉静音段落后，也可以不进行分段，只保留一个语音段落。

[0086] S202，提取语音信息的语音特征向量和语速特征向量，并基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息；该状态信息中包括指示请求者当前是否处于醉酒状态的信息。

[0087] 这里的语音信息即可以指去除掉静音片段后的语音信息，然后提取语音信息中能够表达人类声学特征构成的语音特征向量，以及提取乘客叙述时的语速特征构成的语速特征向量。

[0088] 这里指示请求者当前是否处于醉酒状态的信息，比如可以通过设定的数字进行表示，如通过1001表示醉酒，通过1002表示不醉酒，或者可以通过文字进行表示，即直接表示为“醉酒”或者“不醉酒”。

[0089] S203，基于状态信息，提示服务提供端确认是否接受订单。

[0090] 在确定出请求者当前的状态信息后，可以通过声音提示服务提供端是否接受该订单，或者控制服务提供端显示是否接受该订单的触发按钮，以便司机能够自主进行选择。

[0091] 比如，若确定出乘客当前处于醉酒状态，则可以进行声音提示司机“该乘客处于醉酒状态，请确定是否进行接单”，或者可以在司机的移动终端上进行显示“该乘客处于醉酒状态，请确定是否进行接单”，这样司机能够提前确定乘客的当前状态，从而能够提前采取相应措施以对乘车环境的安全性提前进行风险控制。

[0092] 具体地，在得到乘客的语音信息后，如图3所示，可以按照以下过程提取语音信息的语音特征向量，具体包括步骤S301～S303：

[0093] S301，对语音信息中的各个语音段落分别进行分帧处理，得到每个语音段落对应的语音帧。

[0094] 这里得到语音信息后，对该语音信息中的各个语音段落进行分帧处理，比如按照10ms的间隔进行分帧，将每个语音段落成多个语音帧。

[0095] S302，针对每个语音段落，提取该语音段落中每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差，并基于语音帧特征、语音帧特征差和预先设定的语音段落特征函数，确定该语音段落的第一段落语音特征向量。

[0096] 这里的语音帧特征可以包括语音帧的基频特征、梅尔频率倒谱系数、过零率、谐噪

比和能量等声学特征，这里语音帧与其相邻的语音帧之间的语音帧特征差，在本申请实施例中可以指该语音帧与其前一个语音帧之间的基频特征差、梅尔频率倒谱系数差、过零率差、谐噪比差和能量差等声学特征差。

[0097] 具体地，基频特征、梅尔频率倒谱系数、过零率、谐噪比和能量通过以下方式提取：

[0098] (1)基频(fundamental frequency，F0)是基音的振动频率，它决定语音音调的高低，实际应用中常用最高音pitch来表示基频，这里在提取语音帧的基频特征时，即提取该语音帧的pitch。

[0099] (2)梅尔频率倒谱系数(Mel-frequency cepstral coefficients，MFCC)是模拟人耳听觉感知机理设计的一种特征，其提取方法是：首先对语音帧做短时傅里叶变换(short-term Fourier transform，STFT)得到语谱上的能量分布，接下来将语谱通过一组梅尔谱上均匀分布的三角滤波器，相邻滤波器之间有一半重叠部分，得到每帧在每个梅尔滤波器中的能量水平。最后对三角滤波器组的输出求对数，在通过离散余弦变换(discrete cosine transform，DCT)进行求倒谱的操作，同时也完成了特征的去相关操作。通常只保留前12个DCT系数，因为舍弃高倒频域值的DCT系数可以起到一个类似低通滤波器的作用，能够使信号平滑化，提高语音信号处理的性能。

[0100] 本申请实施例中的梅尔频率倒谱系数MFCC包括12个，则语音帧特征中的梅尔频率倒谱系数有12个。

[0101] (3)过零率(zero crossing rate，ZCR)表示单位采样点内信号通过零点的次数。其计算公式为：

[0102] $$\mathrm{ZCR}_n=\frac{1}{N-1}\sum_{m=0}^{N-1} f(x_n(m)x_n(m+1)<0)$$；

[0103] 其中，$$f(A)=\begin{cases}1 & A\text{是真}\\ 0 & A\text{是假}\end{cases}$$；

[0104] 其中，N表示语音帧的帧长(或者可以指语音帧中采样点的个数，比如若语音帧为间隔10ms采集到的，且每1ms均有一个采样点，则这里的帧长即为10)，$x_n(m)$表示第n帧的第m个信号采样点。过零率常常用来区分清音和浊音，前者过零率较高，后者较低。

[0105] (4)谐噪比(harmonics-to-noise ratio，HNR)是由自相关系数(autocorrelation coefficient function，ACF)计算得到的，可以反映发声概率，其计算公式为：

[0106] $$HNR_n=10\log\frac{ACF(T_0)}{ACF(0)-ACF(T_0)}$$；

[0107] 其中，$$ACF(\tau)=\sum_{m=0}^{N-1-\tau} x(m)x(m+\tau)$$；

[0108] 其中，T_0代表的是基音周期(fundamental period)，τ表示同一个语音帧中相邻两个采样点之间的间隔时间。

[0109] (5)能量：语音信号短时能量的计算方法为一帧内各点信号值的平方和，即：

[0110] $$E_n=\sum_{m=0}^{N-1} x_n^2(m)$$

[0111] 能量反映了音频信号的振幅。通过能量，可以大体判断各帧包含的信息多少，也可

以用来区分浊音段和轻音段、有声段和静音段的分界等。

[0112] 通过以上声学特征，能够表达出请求者语音信息中的声音特点，从而为后续确定请求者的状态信息做参考量。

[0113] 针对语音信息中的任一语音段落，当提取出该任一语音段落中的每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差后，按照设定顺序，得到每个语音帧对应的语音帧特征向量，比如，针对该任一语音段落中的每个语音帧，得到的的任一语音帧的语音帧特征向量为包括上述语音帧特征和语音帧特征差组成的特征向量，以第n个语音帧为例，其得到的语音帧特征向量可以通过以下向量表示：

[0114] $Y_n=(pitch_n\ EMCC_{n1}\ \ldots\ EMCC_{n12}\ \ldots\ E_n\ \Delta pitch_n\ \Delta EMCC_{n1}\ \ldots\ \Delta EMCC_{n12}\ \ldots\ \Delta E_n)^T$；

[0115] 这里第n个语言帧的语音特征向量中，第一个省略号处为$EMCC_{n2}$～$EMCC_{n11}$，第二个省略号处依次为ZCR_n和HNR_n，第三个省略号处为$\Delta EMCC_{n2}$～$\Delta EMCC_{n11}$，第四个省略号处依次为ΔZCR_n和ΔHNR_n。

[0116] 可以看到第n个语音帧的语音帧特征向量是32维度的特征向量，这里的n可以指该任一语音段落中的任一语音帧，若该任一语音段落包括10个语音帧，在得到该任一语音段落中10个语音帧对应的语音帧特征向量为10个语音帧特征向量(Y_1～Y_{10})后，针对这10个语音帧特征向量中的属于同一维度的语音帧特征或语音帧特征差按照预先设定的语音段落特征函数进行处理，确定该任一语音段落对应的第一段落语音特征向量，这里的同一维度语音帧特征向量中的元素位置，比如一个向量包括32个元素，即这里定位为该向量为32维度的向量，属于同一维度是指在每个向量中属于同一个元素位置的元素。

[0117] 具体地，比如针对该任一语音段落包含的所有语音帧特征向量中第一个维度的语音帧特征，即pitch特征，对10个语音帧特征向量中的所有pitch特征分别通过平均值函数、标准差函数，峰度函数，偏度函数、最大值函数，最小值函数，相对最大位置函数、相对最小位置函数，范围函数、以及线性回归函数进行处理，从而得到上述10个语音帧特征向量中所有第一个维度的pitch特征的平均值、标准差，峰度，偏度、最大值，最小值，相对最大位置、相对最小位置、范围、以及线性回归函数的抵消项、斜率和最小均方差这12个函数值，即针对该任一语音段落中每个语音帧特征向量中属于同一维度的语音帧特征或语音帧特征差按照上述函数计算每个维度对应的12个函数值。

[0118] 按照上述方式，即能够得到该任一语音段落的所有语音帧特征和语音帧特征差对应的函数值，然后按照设定的语音帧特征和语音帧特征差顺序以及函数值顺序，组成该任一语音段落的第一段落语音特征向量，若每个语音帧特征向量为32维度时，语音帧特征和语音帧特征差分别对应的函数值个数为12个时，即得到的该第一段落语音特征向量即可以为32*12即384维度的特征向量。

[0119] 为了便于描述，将任一语音段落中每个语音帧特征向量中属于同一维度语音帧特征或语音帧特征差作为一个样本组，比如上述任一语音段落中包括10个语音帧特征向量，每个语音帧特征向量均为32维度，若针对属于第一维度的pitch特征，则该pitch特征样本组总共包括10个样本，32个维度则包括32个样本组，下面分别介绍平均值、标准差，峰度，偏度、最大值，最小值，相对最大位置、相对最小位置、范围、抵消项、斜率和最小均方差的含义：

[0120] 平均值即表示任一语音段落中所有语音帧特征向量中每个样本组中所有样本的平均值，比如针对pitch特征样本组，这里的平均值即该任一语音段落中所有语音帧特征向量中的pitch的平均值。

[0121] 标准差即表示任一语音段落中所有语音帧特征向量中每组样本的标准差。

[0122] 峰度(kurtosis)是描述样本分布的形态与正态分布相比的陡缓程度的统计量，峰度函数通过以下公式表示：

[0123] $$\text{Kurtosis}=\frac{1}{(n-1)D^4}\sum_{i=1}^{n}(x_i-\bar{x})^4-3;$$

[0124] 其中，D为属于该任一语音段落中所有语音帧特征向量中每个样本组的样本方差，$\bar{x}$表示每个样本组的样本均值。

[0125] 偏度(Skewness)与峰度类似，同样是一种描述样本分布形态的统计值，其描述的是某总体的样本分布的对称性，偏度的计算公式如下：

[0126] $$\text{Skewness}=\frac{1}{(n-1)D^3}\sum_{i=1}^{n}(x_i-\bar{x})^3;$$

[0127] 最大值和最小值分别指属于同一个样本组中的最大值和最小值。

[0128] 相对最大位置指的是同一个样本组中最大值所属于的语音帧在该任一语音段落中的位置、相对最小位置指的是同一样本中最小值所属于的语音帧在该任一语音段落中的位置，比如针对pitch特征组成的样本中，若发现最大的pitch来自于该任一语音段落中的第3个位置的语音帧，则相对最大位置为3，若发现最小的pitch来自于该任一语音段落中的第7个位置的语音帧，则相对最小位置为7。

[0129] 范围(Range)指的是同一个样本组中，最大值与最小值之差:Range＝Max-Min;其中，Max是最大值，Min是最小值。

[0130] 抵消项、斜率和最小均方差分别指将同一个样本组构成线性回归函数后，该线性回归函数对应的截距、斜率和最小均方差。

[0131] 按照上述方法，即可以确定出语音信息中每个语音段落的第一段落语音特征向量，若该语音信息包括10个语音段落，即可以得到10个384维度的第一段落语音特征向量。

[0132] 这里将语音帧特征向量转换为第一段落语音特征向量，即得到语音信息中各个语音段落的特征向量，这样能够完整的表达请求者的语音段落的声学特征，便于后期确定请求者的状态信息。

[0133] 当然，以上给出的预先设定的语音段落特征函数并不限于本申请给出的上述几种，上述几种语言段落特征函数仅仅为一种具体实施例。

[0134] S303，基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量。

[0135] 考虑到每个人的醉酒状态大多是不同的，是具有差异性的，具体地，步骤S303中，基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量，如图4所示，包括以下步骤S401～S403：

[0136] S401，针对每个语音段落，基于该语音段落的第一段落语音特征向量和预先存储的清醒状态语音特征向量，确定每个语音段落的差异性语音特征向量。

[0137] 这里的清醒状态语音特征向量即预先采集的大量处于清醒状态的乘客的语音信

息，并对每个处于清醒状态的乘客的语音信息按照上述方式处理，得到每个处于清醒状态的乘客的第一段落语音特征向量，然后对这些乘客的所有第一段落语音特征向量进行取平均计算，得到平均的第一段落语音特征信息，并将该平均的第一段落语音信息作为清醒状态语音特征向量进行存储。

[0138] 这里针对每个语音段落的第一段落语音特征向量，将该第一段落语音特征向量与请求状态语音特征向量进行求差并取绝对值，即得到每个语音段落的差异性语音特征向量，若以C_i表示差异性语音特征向量中第i个元素位置的差异性特征，以D_i表示第一段落语音特征向量中第i个元素位置的第一段落语音特征，以Q_i表示清醒状态语音特征向量中第i个元素位置的清醒状态语音特征，则可以通过以下公式确定差异性语音特征向量中的每个差异性语音特征：

[0139] $C_i=|D_i-Q_i|$；

[0140] 其中，若差异性$i\in(1,K)$，K表示差异性特征向量、第一段落语音特征向量或者清醒状态语音特征向量的维度，这里i从1到K依次取值，这样即可以得到差异性特征向量中每个元素位置的值，即每个元素位置的差异性特征。

[0141] 若第一段落语音特征向量为上述提到的384维度的特征向量，这里的差异性语音特征向量也为384维度的特征向量，即差异性语音特征向量的维度与第一段落语音特征向量的维度相同。

[0142] S402，基于每个语音段落的第一段落语音特征向量和差异性语音特征向量，确定每个语音段落的第二段落语音特征向量。

[0143] 在得到每个语音段落的差异性语音特征向量后，将每个语音段落的差异性语音特征向量与该语音段落的第一段落语音特征向量进行拼接，即得到各个语音段落的第二段落语音特征向量，比如第一段落语音特征向量为384维度的特征向量，差异性语音特征向量也为384维度的特征向量，则第二段落语音特征向量记为768维度的特征向量。

[0144] S403，对各个第二段落语音特征向量进行合并，得到语音信息的语音特征向量。

[0145] 在得到语音信息中各个语音段落的第二段落语音特征向量后，将这些第二段落语音特征向量进行合并，比如本申请实施例语音信息中包括10个语音段落，即得到10个768维度的第二段落语音特征向量，即这10个768维度的第二段落语音特征向量即为该语音信息的语音特征向量。

[0146] 这里的第二段落语音特征向量即可以表示请求者与其他乘客不同的段落语音特征向量，具有一定的差异性，通过第二段落语音特征向量能够更准确地确定请求者的状态信息。

[0147] 另外，在得到乘客的语音信息后，如图5所示，可以按照以下过程提取语音信息的语速特征向量，具体包括以下步骤S501～S504：

[0148] S501，将语音信息中的各个语音段落转换为文本段落，每个文本段落包括多个字符。

[0149] S502，基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长，确定每个文本段落的语速。

[0150] 比如将语音信息中任一语音段落转换为文本段落后，得到该任一语音段落对应的文本段落，该文本段落可以包括M个字符，若该任一语音段落的时长为N，则该任一文本段落

的语速V可以通过V＝M/N确定。

[0151] S503，基于每个文本段落对应的语速，确定语音信息的最大语速、最小语速和平均语速。

[0152] S504，基于语音信息的最大语速、最小语速和平均语速，提取语音信息的语速特征向量。

[0153] 按照上述方式，可以计算出语音信息中每个文本段落的语速，然后在这些语速中确定出最大语速、最小语速和平均语速，然后将该语音信息的最大语速、最小语速和平均语速进行组合，即得到一个3维度的语速特征向量。

[0154] 这里因为醉酒状态能够让乘客的语速变得缓慢，说话吞吞，故本申请实施例中将语速特征向量作为确定请求者状态信息的参考量，提高了识别请求者的状态信息的准确性。

[0155] 在按照上述方式提取出语音信息的语音特征向量和语速特征向量后，就可以基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息，如图6所示，具体包括以下步骤S601～S602：

[0156] S601，基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量。

[0157] 这里，第一分值特征向量包括语音信息中每个语音段落指示醉酒状态的概率值，第二分值特征向量包括语音信息中每个语音段落指示不醉酒状态的概率值。

[0158] 具体地，这里的语音特征向量中包括多个第二段落语音特征向量，这里第一分值特征向量中每个分值特征即表示语言特征向量中对应的第二段落语音特征向量为醉酒状态的概率值，第二分值特征向量中每个分值特征即表示语言特征向量中对应的第二段落语音特征向量为不醉酒状态的概率值。

[0159] 具体地，这里基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量，包括：

[0160] 将语音特征向量输入预先训练的声音识别模型中的段级别分类器，得到语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量。

[0161] 这里的声音识别模型以及声音识别模型中的段级别分类器将在后文进行介绍，这里将语音特征向量输入该声音识别模型中的段级别分类器后，即可以得到由语音特征向量中每个第二段落语音特征向量属于醉酒状态的概率值组成的第一分值特征向量，以及得到由每个第二段落语音特征向量属于不醉酒状态的概率值组成的第二分值特征向量。

[0162] S602，基于第一分值特征向量、第二分值特征向量以及语速特征向量，确定请求者当前的状态信息。

[0163] 在得到第一分值特征向量和第二分值特征向量后，可以结合上文提到的语速特征向量，确定出请求者当前的状态信息，该过程具体包括：

[0164] (1)基于第一分值特征向量、第二分值特征向量和预先设定的语音分值特征函数，确定语音信息的分值特征向量；

[0165] (2)将分值特征向量和语速特征向量合并后，输入声音识别模型中的状态级别分类器，确定请求者当前的状态信息。

[0166] 这里，在得到第一分值特征向量和第二分值特征向量后，可以通过预先设定的语

音分值特征函数，得到能够表示语音信息的分值特征向量，比如通过最大值函数、最小值函数、平均值函数、以及分位数函数，分别计算第一分值特征向量和第二分值特征向量中各个分值特征的最大值、最小值、平均值以及9分位数，即得到与第一分值特向量相关的12个函数值，以及与第二分值特向量相关的12个函数值，然后将这24个函数值进行拼接，得到24维度的语音信息的分值特征向量。

[0167] 比如，针对第一分值特征向量，这里的9分位数包括对该第一分值特征向量中所有分值特征进行一分位时对应的分值特征、二分位时对应的分值特征、三分位时对应的分值特征、四分位时对应的分值特征、五分位时对应的分值特征、六分位时对应的分值特征、七分位时对应的分值特征、八分位时对应的分值特征以及九分位时对应的分值特征，即9个分值特征。

[0168] 然后，再将语音信息的分值特征向量与语速特征向量进行合并后，可以得到27维度的特征向量，即包括24维度的分值特征向量和3维度的语速特征向量，将合并后的分值特征向量和语速特征向量输入声音识别模型中的状态级别分类器，即可确定请求者当前的状态信息。

[0169] 或者，这里可以直接将第一分值特征向量、第二分值特征向量和语速特征向量输入声音识别模型中的状态级别分类器，由声音识别模型中的状态级别分类器基于第一分值特征向量、第二分值特征向量和预先设定的语音分值特征函数，确定语音信息的分值特征向量，然后再将分值特征向量和语速特征向量合并后，确定请求者当前的状态信息。

[0170] 本申请实施例中，通过语音特征向量和语速特征向量共同确定请求者当前的状态信息，一方面考虑请求者当前的声学特征，另一方面考虑请求者当前的语速特征，两种特征共同确定请求者的状态信息，能够更加准确地确定该请求者当前是否处于醉酒状态。

[0171] 以下将针对上述提到的声音识别模型进行介绍，按照以下方式训练声音识别模型：

[0172] (1)构建声音识别模型的段级别分类器和状态级别分类器。

[0173] (2)获取预先构建的训练样本库，训练样本库包括多个训练语音信息对应的训练语音特征向量、训练语速特征向量以及每个训练语音信息对应的状态信息。

[0174] 这里训练样本库中可以包括大量乘客的训练语音信息，这些乘客的状态信息是确定的，比如可以包括1000个处于醉酒状态的乘客和1000个处于不醉酒状态的乘客，这样即能够得到2000个训练语音信息对应的训练语音特征向量和训练语速特征向量。

[0175] 这里的每个训练语音特征向量可以由至少一个第二训练段落语音特征向量，比如每个训练语音特征向量由10个第二训练段落语音特征向量，这里的第二训练段落语音特征向量的确定过程与上文提到的确定第二段落语音特征向量的过程相似，在此不再赘述。

[0176] 另外，这里每个训练语音信息的训练语速特征向量与上文提到的确定语音信息的语速特征向量的过程相似，在此不再赘述。

[0177] (3)依次将每个训练语音信息的训练语音特征向量输入段级别分类器，得到该训练语音信息对应的第一训练分值特征向量和第二训练分值特征向量；将第一训练分值特征向量、第二训练分值特征向量和训练语速特征向量作为状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为声音识别模型的输出变量，训练得到声音识别模型的模型参数信息。

[0178] 本申请实施例中的声音识别模型可以包括两种分类器，第一种即段级别分类器，其输入端可以为训练语音特征向量，输出即为训练语音特征向量中每个第二训练段落语音特征向量属于醉酒状态的概率值组成的第一训练分值特征向量，以及每个第二训练段落语音特征向量属于不醉酒状态的概率值组成的第二训练分值特征向量。

[0179] 然后可以将第一训练分值特征向量、第二训练分值特征向量和语速特征向量作为状态级别分类器的输入变量，具体地，可以先基于第一训练分值特征向量、第二训练分值特征向量和预先设定的语音分值特征函数确定训练语音信息的训练分值特征向量，这里确定训练分值特征向量的过程与上文提到的确定语音信息的分值特征向量的过程相似，在此不再赘述，然后同样将训练分值特征向量和训练语速特征向量合并后再作为状态级别分类器的输入变量，这里的输入变量均携带有其属于的乘客的标码信息，这里的编码信息即可以指示该合并后的训练分值特征向量和训练语速特征向量来自哪个乘客，然后将该训练语音信息对应的状态信息作为声音识别模型的输出变量，训练得到声音识别模型的模型参数信息。

[0180] 在具体训练过程中，声音识别模型中的段级别分类器和状态级别分类器可以分开进行训练，比如先训练段级分类器中的模型参数信息，可以将各个语音信息的训练语音特征向量输入段级分类器，得到第一训练分值特征向量和第二训练分值特征向量，然后将第一训练分值特征向量和第二训练分值特征向量输入预设的损失函数，通过调整段级分类器中的模型参数信息，直至损失函数收敛，即得到段级分类器中的模型参数信息。

[0181] 然后将得到的第一训练分值特征向量、第二训练分值特征向量和训练语速特征向量作为状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为声音识别模型的输出变量，训练状态级别分类器的模型参数信息。

[0182] 在得到模型参数信息后，为了验证声音识别模型的测试性能，一般还需要通过测试样本对上述得到的声音识别模型进行测试，具体过程如下：

[0183] (1)获取预先构建的测试样本库，测试样本库包括多个测试语音信息对应的测试语音特征向量、测试语速特征向量以及每个测试语音信息对应的真实状态信息。

[0184] 这里测试样本库可以包括大量已知状态信息的乘客的测试语音信息，这里的测试语音特征向量和测试语速特征向量与上文的语音特征向量和语速特征向量的确定过程相似，在此不再赘述。

[0185] (2)依次将测试样本库中的每个测试语音特征向量输入声音识别模型的段级分类器，得到测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量；

[0186] 这里将每个测试语音特征向量输入声音识别模型的段级分类器，可以得到测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量，这里的测试语音特征向量包含乘客的编码信息。

[0187] (3)将测试样本库中每个测试语音对应的第一测试分值特征向量、第二测试分值特征向量和测试语速特征向量输入声音识别模型的状态级别分类器，得到测试样本库中的每个测试语音对应的测试状态信息；

[0188] 这里同样可以先基于第一测试分值特征向量、第二测试分值特征向量和预先设定的语音分值特征函数确定测试语音信息的测试分值特征向量，然后再将测试分值特征向量

和测试语速特征向量合并后输入声音识别模型的状态级别分类器，得到测试样本库中的每个测试语音对应的测试状态信息。

[0189] (4)基于真实状态信息和测试状态信息，确定声音识别模型的精确率和召回率。

[0190] 具体地，本申请可以通过精确率和召回率来作为声音识别模型的测试性能评价指标，为了说明精确率和召回率的含义，本申请实施例引入四种分类情况：真正例(True Positives,TP)、假正例(False Positives,FP)、假反例(False Negatives,FN)和真反例(True Negatives,TN)，其具体含义如下表所示：

[0191]

	正类（醉酒）	负类（清醒）
被检索到	True Positives （TP），正类样本判定为正类，即分类结果正确	False Positives （FP），负类样本判定为正类，即其他类别的样本判定为当前类
未被检索到	False Negatives（FN），正类样本判定为负类，即当前类的样本没有被正确分类	True Negatives（TN），负类样本判定为负类

[0192] 精确率Precision计算的是所有被正确分类的样本占所有实际被判别为该类的样本比例，公式为：

[0193] $$\text{Precision}=\frac{TP}{TP+FP};$$

[0194] 召回率Recall计算的是所有被正确分类的样本(TP)占该类实际样本数量的比例，公式为：

[0195] $$\text{Recall}=\frac{TP}{TP+FN};$$

[0196] 这样，按照精确率公式和召回率公式，即可得到声音识别模型的精确率和召回率。

[0197] (5)若精确率和召回率不满足设定条件，更新声音识别模型中的模型训练参数和/或训练样本库，重新训练声音识别模型，直至精确率和召回率满足设定条件。

[0198] 这里的设定条件包括：(1)精确率不小于设定精确率同时召回率不小于设定召回率；(2)精确率不小于设定精确率，召回率不限定；(3)精确率不限定，召回率不小于设定召回率；(4)与精确率和召回率相关的鲁棒性符合设定鲁棒性条件。

[0199] 另外，本申请实施例还可以通过精确率和召回率得到声音识别模型的Precision-Recall曲线，如图7所示，可以从图7中看到，这里的AUC指标为0.819，AP指标为0.122，这里的AUC指标和AP指标能够表示声音识别模型是否满足设定鲁棒性条件。

[0200] 这里的AUC的全称是Area Under Curve，就是ROC曲线和x轴(FPR轴)之间的面积，AP的全称是Average Precision，指的是平均精确率，具体可以为Precision-Recall曲线和x轴和y轴之间的面积。

[0201] 当设定条件确定后，若经过测试样本测试后，确定声音识别模型的精确率和召回率不满足设定条件后，可以更新声音识别模型中的模型训练参数或训练样本库中的训练样本重新训练声音识别模型，或者同时更新声音识别模型中的模型训练参数以及训练样本库中的训练样本重新训练声音识别模型，直至直至精确率和召回率满足设定条件后，停止训练，即得到训练完成的声音识别模型。

[0202] 基于同一申请构思，本申请实施例中还提供了与订单处理方法对应的订单处理装置，由于本申请实施例中的装置解决问题的原理与本申请实施例上述订单处理方法相似，因此装置的实施可以参见方法的实施，重复之处不再赘述。

[0203] 参照图8所示，为本申请提供的一种订单处理装置800的示意图，该订单处理装置800包括：获取模块801、确定模块802、提示模块803；其中，

[0204] 获取模块801，用于在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息，并将语音信息传输至确定模块802；

[0205] 确定模块802，用于提取语音信息的语音特征向量和语速特征向量，并基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息，并将该状态信息传输至提示模块803；状态信息中包括指示请求者当前是否处于醉酒状态的信息；

[0206] 提示模块803，用于基于状态信息，提示服务提供端确认是否接受订单。

[0207] 在一种实施方式中，订单处理装置还包括处理模块804，处理模块804在获取模块801获取服务请求端发送的语音信息之后，确定模块802提取语音信息的语音特征向量和语速特征向量之前，用于：

[0208] 对语音信息进行语音端点检测，得到至少一个语音段落和静音段落；

[0209] 删除语音信息中的静音段落。

[0210] 在一种实施方式中，确定模块802具体用于根据以下步骤提取语音信息时的特征向量：

[0211] 对语音信息中的各个语音段落分别进行分帧处理，得到每个语音段落对应的语音帧；

[0212] 针对每个语音段落，提取该语音段落中每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差，并基于语音帧特征、语音帧特征差和预先设定的语音段落特征函数，确定该语音段落的第一段落语音特征向量；

[0213] 基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量。

[0214] 在一种实施方式中，确定模块802具体用于根据以下步骤基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量：

[0215] 针对每个语音段落，基于该语音段落的第一段落语音特征向量和预先存储的清醒状态语音特征向量，确定每个语音段落的差异性语音特征向量；

[0216] 基于每个语音段落的第一段落语音特征向量和差异性语音特征向量，确定每个语音段落的第二段落语音特征向量；

[0217] 对各个第二段落语音特征向量进行合并，得到语音信息的语音特征向量。

[0218] 在一种实施方式中，确定模块802具体用于根据以下步骤提取语音信息的语速特征向量：

[0219] 将语音信息中的各个语音段落转换为文本段落，每个文本段落包括多个字符；

[0220] 基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长，确定每个文本段落的语速；

[0221] 基于每个文本段落对应的语速，确定语音信息的最大语速、最小语速和平均语速；

[0222] 基于语音信息的最大语速、最小语速和平均语速，提取语音信息的语速特征向量。

[0223] 在一种实施方式中，确定模块802具体用于根据以下步骤基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息：

[0224] 基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量，第一分值特征向量包括语音信息中每个语音段落指示醉酒状态的概率值，第二分值特征向量包括语音信息中每个语音段落指示不醉酒状态的概率值；

[0225] 基于第一分值特征向量、第二分值特征向量以及语速特征向量，确定请求者当前的状态信息。

[0226] 在一种实施方式中，确定模块802具体用于根据以下步骤基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量：

[0227] 将语音特征向量输入预先训练的声音识别模型中的段级别分类器，得到语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量；

[0228] 基于第一分值特征向量、第二分值特征向量和预先设定的语音分值特征函数，确定语音信息的分值特征向量；

[0229] 将分值特征向量和语速特征向量合并后，输入声音识别模型中的状态级别分类器，确定请求者当前的状态信息。

[0230] 在一种实施方式中，还包括模型训练模块805，模型训练模块805用于根据以下步骤训练声音识别模型：

[0231] 构建声音识别模型的段级别分类器和状态级别分类器；

[0232] 获取预先构建的训练样本库，训练样本库包括多个训练语音信息对应的训练语音特征向量、训练语速特征向量以及每个训练语音信息对应的状态信息；

[0233] 依次将每个训练语音信息的训练语音特征向量输入段级别分类器，得到该训练语音信息对应的第一训练分值特征向量和第二训练分值特征向量；将第一训练分值特征向量、第二训练分值特征向量和语速特征向量作为状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为声音识别模型的输出变量，训练得到声音识别模型的模型参数信息。

[0234] 在一种实施方式中，订单处理装置还包括模型测试模块806，模型测试模块806用于根据以下步骤测试声音识别模型：

[0235] 获取预先构建的测试样本库，测试样本库包括多个测试语音信息对应的测试语音特征向量、测试语速特征向量以及每个测试语音信息对应的真实状态信息；

[0236] 依次将测试样本库中的每个测试语音特征向量输入声音识别模型的段级分类器，得到测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量；

[0237] 将测试样本库中每个测试语音对应的第一测试分值特征向量、第二测试分值特征向量和测试语速特征向量输入声音识别模型的状态级别分类器，得到测试样本库中的每个测试语音对应的测试状态信息；

[0238] 基于真实状态信息和测试状态信息，确定声音识别模型的精确率和召回率。

[0239] 若精确率小于设定精确率，和/或，召回率小于设定召回率，可以更新声音识别模型中的模型训练参数和训练样本库中的至少一种，使得模型训练模块805重新训练声音识别模型，直至声音识别模型的精确率不小于设定精确率，且召回率不小于设定召回率。

[0240] 本申请实施例还提供了一种电子设备900，如图9所示，为本申请实施例提供的电子设备900的结构示意图，包括：处理器901、存储介质902和总线903，存储介质902存储有处理器901可执行的机器可读指令(比如获取模块801、确定模块802、提示模块803等)，当电子设备900运行时，处理器901与存储介质902之间通过总线903通信，机器可读指令被处理器901执行时执行如下处理：

[0241] 在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息；

[0242] 提取语音信息的语音特征向量和语速特征向量，并基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息；状态信息中包括指示请求者当前是否处于醉酒状态的信息；

[0243] 基于状态信息，提示服务提供端确认是否接受订单。

[0244] 一种可能的实施方式中，在获取服务请求端发送的语音信息之后，提取语音信息的语音特征向量和语速特征向量之前，处理器901执行的指令还包括：

[0245] 对语音信息进行语音端点检测，得到至少一个语音段落和静音段落；

[0246] 删除语音信息中的静音段落。

[0247] 一种可能的实施方式中，处理器901执行的指令包括：

[0248] 对语音信息中的各个语音段落分别进行分帧处理，得到每个语音段落对应的语音帧；

[0249] 针对每个语音段落，提取该语音段落中每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差，并基于语音帧特征、语音帧特征差和预先设定的语音段落特征函数，确定该语音段落的第一段落语音特征向量；

[0250] 基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量。

[0251] 一种可能的实施方式中，处理器901执行的指令包括：

[0252] 针对每个语音段落，基于该语音段落的第一段落语音特征向量和预先存储的清醒状态语音特征向量，确定每个语音段落的差异性语音特征向量；

[0253] 基于每个语音段落的第一段落语音特征向量和差异性语音特征向量，确定每个语音段落的第二段落语音特征向量；

[0254] 对各个第二段落语音特征向量进行合并，得到语音信息的语音特征向量。

[0255] 一种可能的实施方式中，处理器901执行的指令包括：

[0256] 将语音信息中的各个语音段落转换为文本段落，每个文本段落包括多个字符；

[0257] 基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长，确定每个文本段落的语速；

[0258] 基于每个文本段落对应的语速，确定语音信息的最大语速、最小语速和平均语速；

[0259] 基于语音信息的最大语速、最小语速和平均语速，提取语音信息的语速特征向量。

[0260] 一种可能的实施方式中，处理器901执行的指令包括：

[0261] 基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量，第一分值特征向量包括语音信息中每个语音段落指示醉酒状态的概率值，第二分值特征向量包括语音信息中每个语音段落指示不醉酒状态的概率值；

[0262] 基于第一分值特征向量、第二分值特征向量以及语速特征向量，确定请求者当前的状态信息。

[0263] 一种可能的实施方式中，处理器901执行的指令包括：

[0264] 将语音特征向量输入预先训练的声音识别模型中的段级别分类器，得到语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量；

[0265] 基于第一分值特征向量、第二分值特征向量和预先设定的语音分值特征函数，确定语音信息的分值特征向量；

[0266] 将分值特征向量和语速特征向量合并后，输入声音识别模型中的状态级别分类器，确定请求者当前的状态信息。

[0267] 一种可能的实施方式中，处理器901执行的指令还包括：

[0268] 构建声音识别模型的段级别分类器和状态级别分类器；

[0269] 获取预先构建的训练样本库，训练样本库包括多个训练语音信息对应的训练语音特征向量、训练语速特征向量以及每个训练语音信息对应的状态信息；

[0270] 依次将每个训练语音信息的训练语音特征向量输入段级别分类器，得到该训练语音信息对应的第一训练分值特征向量和第二训练分值特征向量；将第一训练分值特征向量、第二训练分值特征向量和训练语速特征向量作为状态级别分类器的输入变量，将该训练语音信息对应的状态信息作为声音识别模型的输出变量，训练得到声音识别模型的模型参数信息。

[0271] 一种可能的实施方式中，处理器901执行的指令还包括：

[0272] 获取预先构建的测试样本库，测试样本库包括多个测试语音信息对应的测试语音特征向量、测试语速特征向量以及每个测试语音信息对应的真实状态信息；

[0273] 依次将测试样本库中的每个测试语音特征向量输入声音识别模型的段级分类器，得到测试样本库中每个测试语音对应的第一测试分值特征向量和第二测试分值特征向量；

[0274] 将测试样本库中每个测试语音对应的第一测试分值特征向量、第二测试分值特征向量和测试语速特征向量输入声音识别模型的状态级别分类器，得到测试样本库中的每个测试语音对应的测试状态信息；

[0275] 基于真实状态信息和测试状态信息，确定声音识别模型的精确率和召回率；

[0276] 若精确率和召回率不满足设定条件，更新声音识别模型中的模型训练参数和/或训练样本库，重新训练声音识别模型，直至精确率和召回率满足设定条件。

[0277] 本申请实施例还提供了一种计算机可读存储介质，计算机可读存储介质上存储有计算机程序，计算机程序被处理器运行时执行上述订单处理方法的步骤。

[0278] 具体地，该存储介质能够为通用的存储介质，如移动磁盘、硬盘等，该存储介质上的计算机程序被运行时，能够执行上述订单处理方法方法，从而解决无法对乘车环境的安全性提前进行风险控制的问题，能够对乘车环境的安全性提前进行风险控制，进而提高了整体乘车环境的安全性。

[0279] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统和装置的具体工作过程，可以参考方法实施例中的对应过程，本申请中不再赘述。在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。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，又例如，多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0280] 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单元来实现本实施例方案的目的。

[0281] 另外，在本申请各个实施例中的各功能单元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。

[0282] 所述功能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时，可以存储在一个处理器可执行的非易失的计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0283] 以上仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

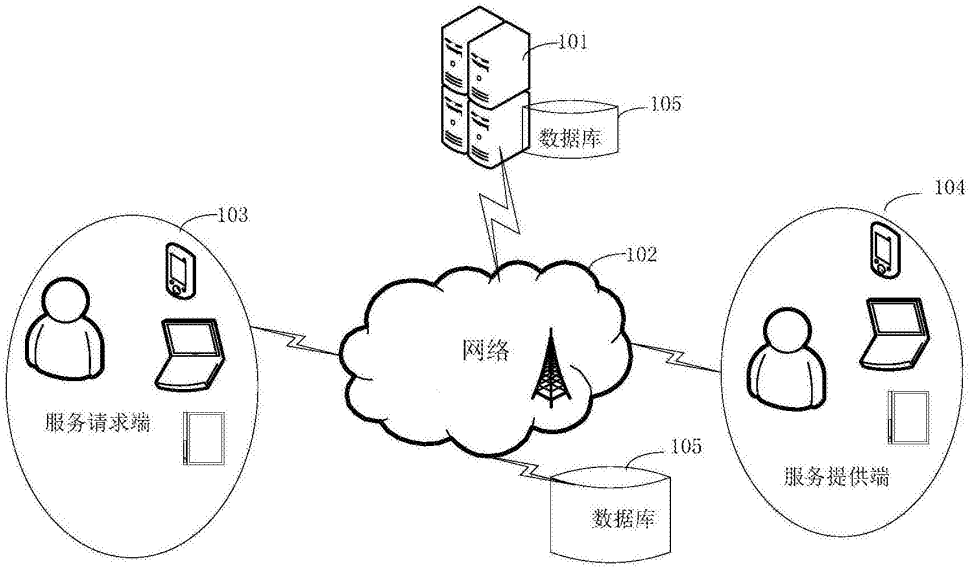

101
105
数据库
104
103
102
网络
服务请求端
105
数据库
服务提供端

图1

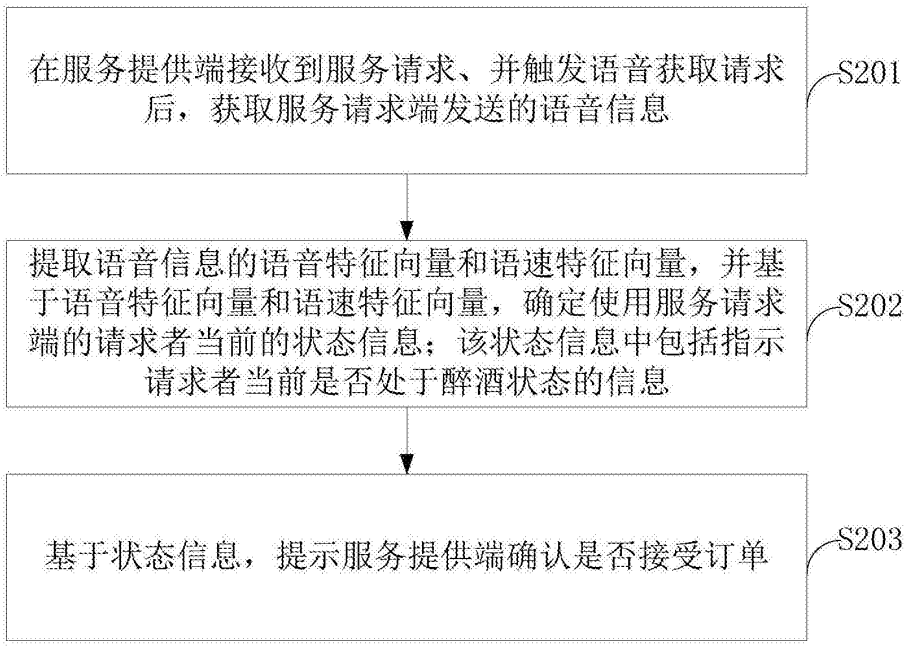

在服务提供端接收到服务请求、并触发语音获取请求后，获取服务请求端发送的语音信息
S201
提取语音信息的语音特征向量和语速特征向量，并基于语音特征向量和语速特征向量，确定使用服务请求端的请求者当前的状态信息；该状态信息中包括指示请求者当前是否处于醉酒状态的信息
S202
基于状态信息，提示服务提供端确认是否接受订单
S203

图2

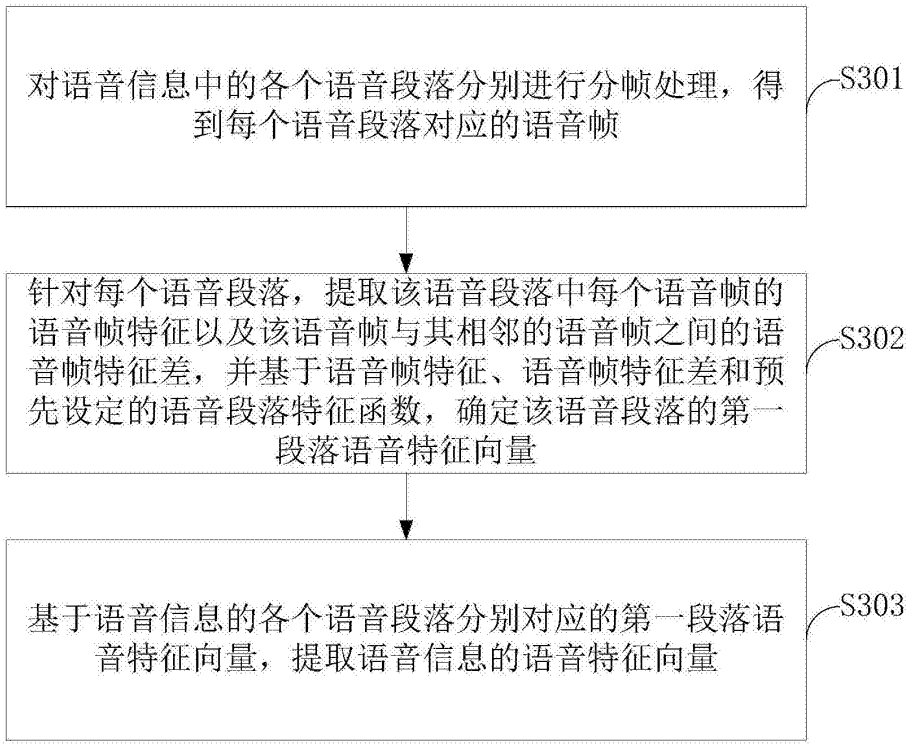

对语音信息中的各个语音段落分别进行分帧处理，得到每个语音段落对应的语音帧
S301
针对每个语音段落，提取该语音段落中每个语音帧的语音帧特征以及该语音帧与其相邻的语音帧之间的语音帧特征差，并基于语音帧特征、语音帧特征差和预先设定的语音段落特征函数，确定该语音段落的第一段落语音特征向量
S302
基于语音信息的各个语音段落分别对应的第一段落语音特征向量，提取语音信息的语音特征向量
S303

图3

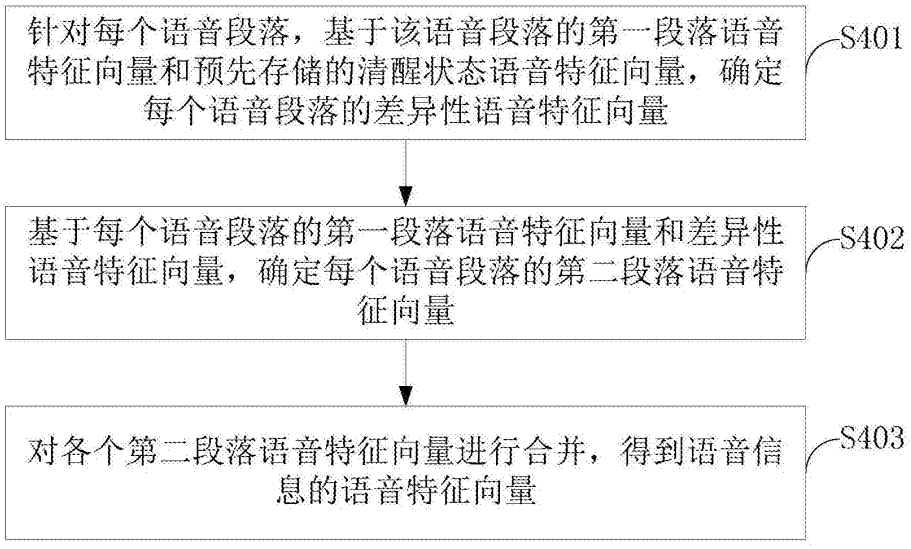

针对每个语音段落，基于该语音段落的第一段落语音特征向量和预先存储的清醒状态语音特征向量，确定每个语音段落的差异性语音特征向量
S401
基于每个语音段落的第一段落语音特征向量和差异性语音特征向量，确定每个语音段落的第二段落语音特征向量
S402
对各个第二段落语音特征向量进行合并，得到语音信息的语音特征向量
S403

图4

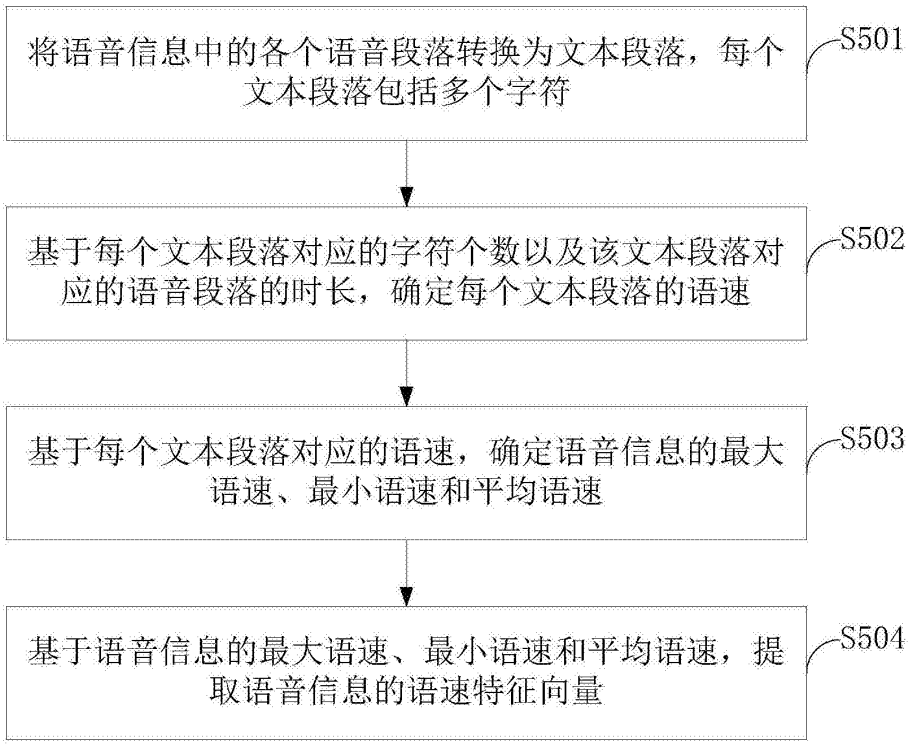

将语音信息中的各个语音段落转换为文本段落，每个文本段落包括多个字符
S501
基于每个文本段落对应的字符个数以及该文本段落对应的语音段落的时长，确定每个文本段落的语速
S502
基于每个文本段落对应的语速，确定语音信息的最大语速、最小语速和平均语速
S503
基于语音信息的最大语速、最小语速和平均语速，提取语音信息的语速特征向量
S504

图5

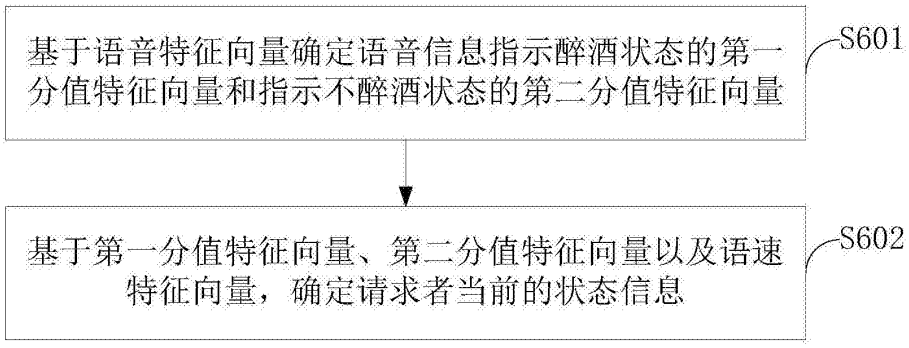

基于语音特征向量确定语音信息指示醉酒状态的第一分值特征向量和指示不醉酒状态的第二分值特征向量
S601
基于第一分值特征向量、第二分值特征向量以及语速特征向量，确定请求者当前的状态信息
S602

图6

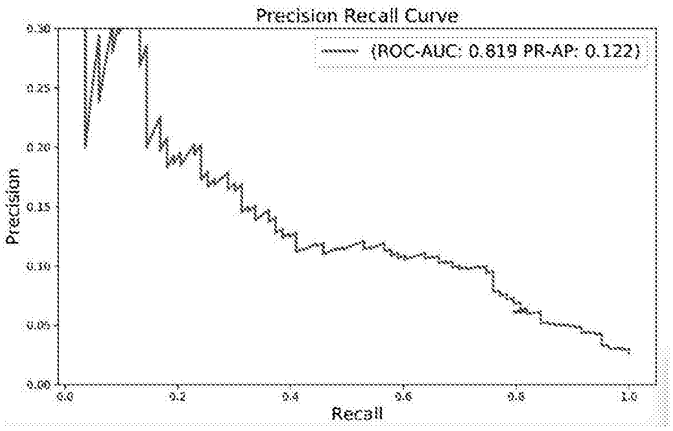

Precision Recall Curve
(ROC-AUC: 0.819 PR-AP: 0.122)
Precision
Recall
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.2
0.4
0.6
0.8
1.0

图7

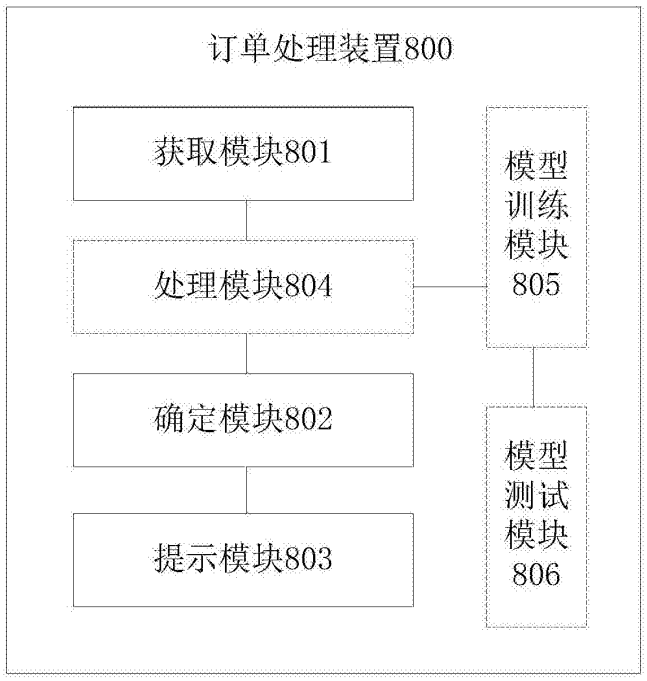

订单处理装置800
获取模块801
处理模块804
确定模块802
提示模块803
模型训练模块805
模型测试模块806

图8

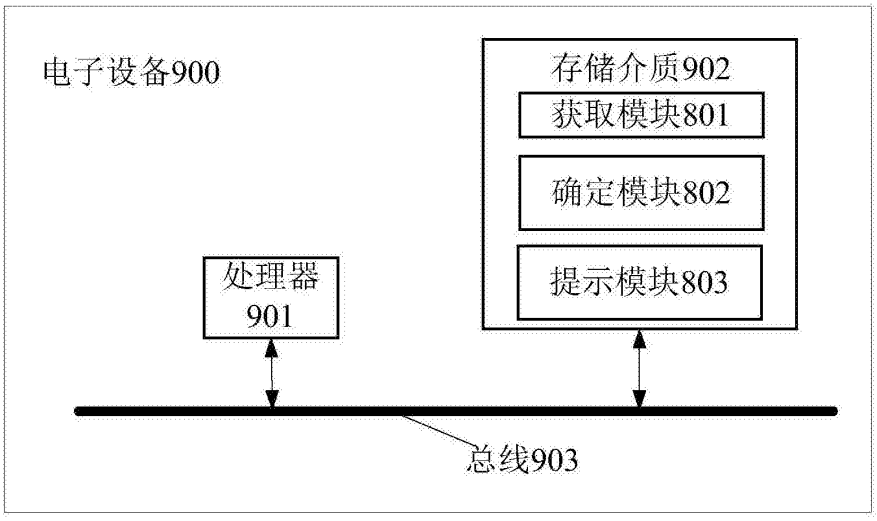


图9