



(12) 发明专利

(10) 授权公告号 CN 102231278 B

(45) 授权公告日 2013.08.21

(21) 申请号 201110156209.3

(22) 申请日 2011.06.10

(73) 专利权人 安徽科大讯飞信息科技股份有限公司

地址 230088 安徽省合肥市高新开发区黄山路616号

(72) 发明人 陈志刚 蒋成林 俞健 魏思 胡郁 胡国平 王智国 刘庆峰

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 赵景平 逯长明

(51) Int. Cl.

G10L 15/26(2006.01)

G10L 15/19(2013.01)

(56) 对比文件

US 2004/0138881 A1, 2004.07.15, 摘要, 说明书第 [0033]-[0035], [0058]-[0099] 段, 图 1, 图 3-9.

CN 86101508 A, 1988.01.06, 摘要、说明书第 17 页倒数第 2 行至第 18 页第 11 行、第 21 页第 3 段至第 23 页倒数第 3 行, 图 9.

CN 1235312 A, 1999.11.17, 全文.

US 2010/0241429 A1, 2010.09.23, 全文.

陆洋海等. 基于知识的语音识别与理解.《中国科学技术大学学报》.1990, 第 20 卷(第 1 期), 43-48.

审查员 万济萍

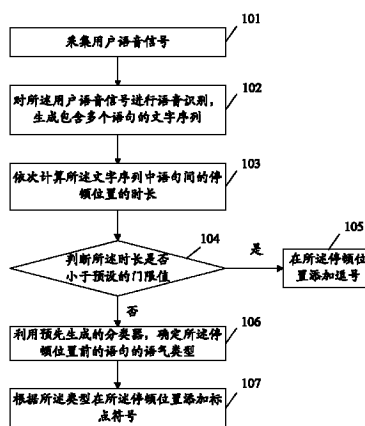
权利要求书2页 说明书8页 附图4页

(54) 发明名称

实现语音识别中自动添加标点符号的方法及系统

(57) 摘要

本发明涉及语音识别技术领域,公开了一种实现语音识别中自动添加标点符号的方法及系统,该方法包括:采集用户语音信号;对所述用户语音信号进行语音识别,生成包含多个语句的文字序列;依次计算所述文字序列中语句间的停顿位置的时长;如果所述时长小于预设的门限值,则在该停顿位置添加逗号;如果所述时长大于或等于所述门限值,则利用预先生成的分类器,确定所述停顿位置前的语句的语气类型,并根据所述类型在所述停顿位置添加标点符号。利用本发明,可以简单方便地实现标点符号的自动添加,并提高标点符号添加的准确性和灵活性。



1. 一种实现语音识别中自动添加标点符号的方法,其特征在于,包括:

采集用户语音信号;

对所述用户语音信号进行语音识别,生成包含多个语句的文字序列;

依次计算所述文字序列中语句间的停顿位置的时长;

如果所述时长小于预设的门限值,则在所述停顿位置添加逗号;

如果所述时长大于或等于所述门限值,则利用预先生成的分类器,确定所述停顿位置前的语句的语气类型,并根据所述类型在所述停顿位置添加标点符号,其中,预先收集文本语料,根据所述文本语料通过训练方式生成文本分类器和/或声学分类器;

所述利用预先生成的分类器,确定所述停顿位置前的语句的语气类型包括:对当前需要识别的语句提取关键词,利用文本分类器计算需要识别的语句属于每个类别的条件概率,和/或根据声学分类器对上述语句进行语气分析,包括韵律分析和频谱分析,根据计算得到的条件概率和/或语气分析结果,确定需要识别的语句的语气类型;

所述根据所述文本语料通过训练方式生成文本分类器包括:从所述文本语料中提取关键词,根据提取的关键词训练生成文本分类器;其中,从所述文本语料中提取关键词具体为:针对句子标点符号为句号、问号、感叹号的语句,分别统计其中出现频率较高、而在其它标点对应的文本中出现较少的词语;根据提取的关键词训练生成文本分类器具体为:将提取的关键词以类型为句号、感叹号、问号结尾的文本中出现的关键词组成一个矢量,每种语气仅保留前N个最有区分性的词语,共组成长度为M的矢量,在前面得到的M个元素的基础上,对于测试训练集中的所有文本语料,分析每个句子中各个关键词是否出现,如果出现,在上述M维矢量中对应的位置填1,其余的位置设为0,然后利用以这些矢量和对应的标点符号为类别,以最大熵为准则,训练出文本分类器;

所述根据所述文本语料通过训练方式生成声学分类器包括:获取不同的录音人对所述文本语料录音生成的语音文件,确定所述语音文件中的发音语气,以所述发音语气为分类目标训练生成声学分类器;其中,确定所述语音文件中的发音语气具体为:通过对所述语音文件进行语气分析,确定所述语音文件中的发音语气,所述语气分析包括韵律分析和频谱分析,韵律分析包括基频提取和规整、一阶、二阶基频差分,频谱分析包括计算平均频率幅度、共振峰参数提取;以所述发音语气为分类目标训练生成声学分类器具体为:对于每一条语句,对于初静寂段以外每一个音节计算平均频谱幅度、基频,加上各维特征的一阶、二阶差分,组成一个矢量,以各语音文件中的发音语气为分类目标训练隐支持向量机分类器,作为声学分类器。

2. 根据权利要求1所述的方法,其特征在于,所述类型包括:陈述句、疑问句和感叹句;

所述根据所述类型在所述停顿位置添加标点符号包括:

如果所述类型为陈述句,则在所述停顿位置添加句号;

如果所述类型为疑问句,则在所述停顿位置添加问号;

如果所述类型为感叹句,则在所述停顿位置添加感叹号。

3. 一种实现语音识别中自动添加标点符号的系统,其特征在于,包括:语料收集单元,用于收集文本语料;

分类器生成单元,用于根据所述文本语料通过训练方式生成文本分类器和/或声学分

类器，

其中，从所述文本语料中提取关键词，根据提取的关键词训练生成文本分类器，其中，从所述文本语料中提取关键词为：针对句子标点符号为句号、问号、感叹号的语句，分别统计其中出现频率较高、而在其它标点对应的文本中出现较少的词语，根据提取的关键词训练生成文本分类器为：将提取的关键词以类型为句号、感叹号、问号结尾的文本中出现的关键词组成一个矢量，每种语气仅保留前 N 个最有区分性的词语，共组成长度为 M 的矢量，在前面得到的 M 个元素的基础上，对于测试训练集中的所有文本语料，分析每个句子中各个关键词是否出现，如果出现，在上述 M 维矢量中对应的位置填 1，其余的位置设为 0，然后利用以这些矢量和对应的标点符号为类别，以最大熵为准则，训练出文本分类器；

获取不同的录音人对所述文本语料录音生成的语音文件，确定所述语音文件中的发音语气，以所述发音语气为分类目标训练生成声学分类器，其中，确定所述语音文件中的发音语气为：通过对所述语音文件进行语气分析，确定所述语音文件中的发音语气，所述语气分析包括韵律分析和频谱分析，韵律分析包括基频提取和规整、一阶、二阶基频差分，频谱分析包括计算平均频率幅度、共振峰参数提取，以所述发音语气为分类目标训练生成声学分类器为：对于每一条语句，对于初静寂段以外每一个音节计算平均频谱幅度、基频，加上各维特征的一阶、二阶差分，组成一个矢量，以各语音文件中的发音语气为分类目标训练隐支持向量机分类器，作为声学分类器；

信号采集单元，用于采集用户语音信号；

语音识别单元，用于对所述用户语音信号进行语音识别，生成包含多个语句的文字序列；

计算单元，用于依次计算所述文字序列中语句间的停顿位置的时长；

判断单元，用于判断所述时长是否小于预设的门限值；

第一标点符号添加单元，用于在所述判断单元判断所述时长小于预设的门限值，在所述停顿位置添加逗号；

语句类型确定单元，用于在所述判断单元判断所述时长大于或等于所述门限值，利用预先生成的文本分类器和 / 或声学分类器，确定所述停顿位置前的语句的语气类型，其中，所述利用预先生成的文本分类器和 / 或声学分类器，确定所述停顿位置前的语句的语气类型包括：对当前需要识别的语句提取关键词，利用文本分类器计算需要识别的语句属于每个类别的条件概率，和 / 或根据声学分类器对上述语句进行语气分析，包括韵律分析和频谱分析，根据计算得到的条件概率和 / 或语气分析结果，确定需要识别的语句的语气类型；

第二标点符号添加单元，用于根据所述语句类型确定单元确定的语句类型在所述停顿位置添加标点符号。

4. 根据权利要求 3 所述的系统，其特征在于，所述类型包括：陈述句、疑问句和感叹句；

所述第二标点符号添加单元，具体用于在所述语句类型确定单元确定的语句类型为陈述句时，在所述停顿位置添加句号；为疑问句时，在所述停顿位置添加问号；为感叹句时，在所述停顿位置添加感叹号。

实现语音识别中自动添加标点符号的方法及系统

技术领域

[0001] 本发明涉及语音识别技术领域,具体涉及一种实现语音识别中自动添加标点符号的方法及系统。

背景技术

[0002] 目前,语音识别系统大多采用基于统计模式识别的方法,首先将语音输入的时域声波转化为一种数字化的矢量特征来描述区分不同的发音,基于该声音特征对所有的发音建立一个声学模型;同时,对于大词表连续语音识别系统来说,需要一个语言模型,该模型中包括所识别语言中的常用文字的使用方法。一般连续语音识别系统的工作过程可以描述为,在巨大的字、词、短语或句子的空间中,找出与给定输入声音特征序列相匹配具有最大概率的字、词、短语或句子,作为识别结果,同时可采用上下文产生装置利用上述识别结果修改当前上下文,便于下一个语音采样所用。

[0003] 由于语音可以通过语气的强弱、语音的停顿以及语调的升降变换来变现平稳、感叹、疑问、喜悦等多种情感,而将语音识别为文字时,就失去了发音的辅助信息,无法将这些情感像运用语言一样表达得那样生动。同时讲话人采用不一样的情感语气进行发音,会使得短语或句子具有不同的意思。例如,“你还没听懂我的意思。”和“你还没听懂我的意思?”这两句话由于语气不同,句意也发生改变,前者是陈述一个事实,后者则为表示疑问的反问句,并包含感到意外的意思,而这两句话从文字上的区别仅是标点符号的不同。可见标点符号在句子中起的作用至关重要,在语音识别中添加标点符号是十分必要的。然而在当前的连续语音识别系统中,对于自动添加标点符号的研究并不多,大多是当语音识别时,中间有停顿的地方识别为逗号,结束时自动添加句号,整个句子都被视为陈述语气,这种方式在某些情形下并不能正确传达出讲话者语义和情感。

[0004] 为此,现有技术中提出了一种利用人在发音过程中产生的噪声,将这些噪声与标点符号相对应,实现在连续语音识别中自动添加标点符号的方案。但是该方案在实际应用中依然存在一定的问题,因为用户的差异性及标点符号的多样性,并不是所有用户都会在讲话中产生足够的噪声,因此这种方案中标点符号的添加缺乏准确性和灵活性。

发明内容

[0005] 本发明实施例针对上述现有技术存在的问题,提供一种实现语音识别中自动添加标点符号的方法及系统,以简单方便地实现标点符号的自动添加,并提高标点符号添加的准确性和灵活性。

[0006] 为此,本发明实施例提供如下技术方案:

[0007] 一种实现语音识别中自动添加标点符号的方法,包括:

[0008] 采集用户语音信号;

[0009] 对所述用户语音信号进行语音识别,生成包含多个语句的文字序列;

[0010] 依次计算所述文字序列中语句间的停顿位置的时长;

- [0011] 如果所述时长小于预设的门限值,则在所述停顿位置添加逗号;
- [0012] 如果所述时长大于或等于所述门限值,则利用预先生成的分类器,确定所述停顿位置前的语句的语气类型,并根据所述类型在所述停顿位置添加标点符号。
- [0013] 一种实现语音识别中自动添加标点符号的系统,包括:
- [0014] 信号采集单元,用于采集用户语音信号;
- [0015] 语音识别单元,用于对所述用户语音信号进行语音识别,生成包含多个语句的文字序列;
- [0016] 计算单元,用于依次计算所述文字序列中语句间的停顿位置的时长;
- [0017] 判断单元,用于判断所述时长是否小于预设的门限值;
- [0018] 第一标点符号添加单元,用于在所述判断单元判断所述时长小于预设的门限值,在所述停顿位置添加逗号;
- [0019] 语句类型确定单元,用于在所述判断单元判断所述时长大于或等于所述门限值,利用预先生成的分类器,确定所述停顿位置前的语句的语气类型;
- [0020] 第二标点符号添加单元,用于根据所述语句类型确定单元确定的语句类型在所述停顿位置添加标点符号。
- [0021] 本发明实施例提供的实现语音识别中自动添加标点符号的方法及系统,通过对采集的用户语音信号进行语音识别,生成包含多个语句的文字序列;依次计算该文字序列中语句间的停顿位置的时长;如果该时长小于预设的门限值,则在该停顿位置添加逗号;如果该时长大于或等于所述门限值,则确定该停顿位置前的语句的语气类型,并根据确定的语句类型在该停顿位置添加标点符号。从而简单高效地实现标点符号的自动添加,并保证了标点符号添加的准确性和灵活性。

附图说明

- [0022] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。
- [0023] 图 1 是本发明实施例实现语音识别中自动添加标点符号的方法的流程图;
- [0024] 图 2 是本发明实施例中训练分类器的流程图;
- [0025] 图 3 是本发明实施例中根据已训练的分类器确定语句的语气类型的流程图;
- [0026] 图 4 是本发明实施例实现语音识别中自动添加标点符号的系统的一种结构示意图;
- [0027] 图 5 是本发明实施例实现语音识别中自动添加标点符号的系统的另一种结构示意图。

具体实施方式

- [0028] 为了使本技术领域的人员更好地理解本发明实施例的方案,下面结合附图和实施方式对本发明实施例作进一步的详细说明。
- [0029] 本发明实施例实现语音识别中自动添加标点符号的方法及系统,通过对采集的用户语音信号进行语音识别,生成包含多个语句的文字序列;依次计算该文字序列中语句间

的停顿位置的时长;如果该时长小于预设的门限值,则在该停顿位置添加逗号;如果该时长大于或等于所述门限值,则确定该停顿位置前的语句的语气类型,并根据确定的语句类型在该停顿位置添加标点符号。从而简单高效地实现标点符号的自动添加,并保证了标点符号添加的准确性和灵活性。

[0030] 如图 1 所示,是本发明实施例实现语音识别中自动添加标点符号的方法的流程图,包括以下步骤:

[0031] 步骤 101,采集用户语音信号。

[0032] 可以利用现有的一些语音信号采集工具进行采集,将用户语音信号表示为语音采样。例如,利用 Windows 语音采集工具采集语音信号,将采集的数据保存 WAV(Wave Audio Files,波形声音文件)格式。当然,还可以采用其他采样格式,比如,PCM(Pulse Code Modulation,脉冲编码调制)、RAW 格式、SND 格式、以及 MP3(Moving Picture Experts Group Audio Layer III) 格式等。

[0033] 步骤 102,对所述用户语音信号进行语音识别,生成包含多个语句的文字序列。

[0034] 在具体应用中,可以根据应用需要,采用不同的语音识别方法,比如,可以采用以下任意一种方法:

[0035] (1)DTW(Dynamic Time Warping,动态时间规整)方法:说话人信息不仅有稳定因素(发声器官的结构和发声习惯),而且有时变因素(语速、语调、重音和韵律)。DTW 将识别模板与参考模板进行时间对比,按照某种距离测度得出两模板间的相似程度。

[0036] (2)VQ(Vector Quantization,矢量量化)方法:将每个人的特定文本训练成码本,识别时将测试文本按此码本进行编码,以量化产生的失真度作为判决标准。

[0037] (3)HMM(Hidden Markov Model,马尔可夫模型)方法:是一种基于转移概率和输出概率的随机模型,它将语音看成由可观察到的符号序列组成的随机过程,符号序列则是发声系统状态序列的输出。在使用隐马尔可夫模型识别时,为每个说话人建立发声模型,通过训练得到状态转移概率矩阵和符号输出概率矩阵。识别时计算未知语音在状态转移过程中的最大概率,根据最大概率对应的模型进行判决。对于与文本无关的说话人识别一般采用各态历经型 HMM;对于与文本有关的说话人识别一般采用从左到右型 HMM。HMM 不需要时间规整,可节约判决时的计算时间和存储量,在目前被广泛应用。其缺点是训练时计算量较大。

[0038] (4)ANN(Artificial Neural Networks,人工神经网络方法):它在某种程度上模拟了生物的感知特性,是一种分布式并行处理结构的网络模型,具有自组织和自学习能力、很强的复杂分类边界区分能力以及对不完全信息的鲁棒性,其性能近似理想的分类器。

[0039] 当然,还可以采用其他方法,具体采用哪种语音识别方法本发明实施例不做限定。而且具体的语音识别过程与现有技术类似,在此不再赘述。

[0040] 经过语音识别处理,可以将采集的用户语音信号生成一个文字序列,在该文字序列中包含了一个或多个语句,如果有多个语句,则在前、后语句间有一定时长的停顿,不同停顿位置的时长可能相同,也可能不同。

[0041] 步骤 103,依次计算所述文字序列中语句间的停顿位置的时长。

[0042] 步骤 104,判断所述时长是否小于预设的门限值;如果是,则执行步骤 105;否则,执行步骤 106。

[0043] 上述门限值可以根据不同的语速来确定,比如在正常的语速下设为 0.2s,如果语速较慢,则设为 0.3s,对此本发明实施例不做限定。

[0044] 步骤 105,在所述停顿位置添加逗号。

[0045] 人在讲话时,为使讲话层次分明,突出重点,吸引听话人的注意力,会根据讲话内容的标点有意识的停顿。通常,在一句话结束时会停顿时间较长,而在一句话之间的停顿时间会较短。因此,如果停顿位置的时长很短,小于设定的门限值,则表明句子没有结束,可以直接进行标点符号的添加,即在该停顿位置添加逗号。

[0046] 步骤 106,利用预先生成的分类器,确定所述停顿位置前的语句的语气类型。

[0047] 如果停顿位置的时长较长,超过了预设的门限值,则确定句子结束。而不管是对汉语还是其他语言,比如英语、日语等语言中,一句话的结束可以有多种不同的标点符号来标识,而且,标点符号具有的含义与人的情感之间有直接的联系,比如,当人在讲话时采用感叹语气时,一般句子结尾为感叹号;讲话人采用疑问语气时,句子结尾为问号;讲话人采用陈述语气时,句子结尾为句号。

[0048] 因此,在本发明实施例中,可以在确定句子结束后,根据停顿位置前的语句的语气类型来添加相应的标点符号,从而保证其正确性。

[0049] 步骤 107,根据所述类型在所述停顿位置添加标点符号。

[0050] 在本发明实施例中,可以将语句的语气类型划分为三种,分别是:陈述句、疑问句和感叹句。相应地,在所述停顿位置添加标点符号时根据该类型添加相应的标点符号,即:如果所述类型为陈述句,则在所述停顿位置添加句号;如果所述类型为疑问句,则在所述停顿位置添加问号;如果所述类型为感叹句,则在所述停顿位置添加感叹号。

[0051] 可见,本发明实施例实现语音识别中自动添加标点符号的方法,通过对采集的用户语音信号进行语音识别,生成包含多个语句的文字序列;依次计算该文字序列中语句间的停顿位置的时长;如果该时长小于预设的门限值,则在该停顿位置添加逗号;如果该时长大于或等于所述门限值,则利用预先生成的分类器,确定该停顿位置前的语句的语气类型,并根据确定的语句类型在该停顿位置添加标点符号。从而简单高效地实现标点符号的自动添加,并保证了标点符号添加的准确性和灵活性。

[0052] 前面提到,在本发明实施例中,如果确定了语句结束,则需要根据语句的语气类型来添加相应的标点符号,具体地,可以利用预先生成的分类器(包括文本分类器和/或声学分类器),确定所述语句的语气类型。下面对其原理及具体实现过程进行详细说明。

[0053] 一般地,文字文本之所以能表达一定的情感和语义,除了与文字本身具有的含义有关,更与文本中采用的语气词和标点符号具有直接的联系。

[0054] 另外,语音信号之所以能够传递情感信息,语气不同能表达不同的含义,是因为其中包含了能反映情感差异的特征,而声学特征在情感信息传递中起到关键作用,语音情感变化主要通过声学特征的差异得到体现。

[0055] 因此,在本发明实施例中,可以采用分别训练文本分类器和声学分类器的方式,当然,也可以采用同时训练文本分类器和声学分类器的方式,利用所述文本分类器和声学分类器对用户语音进行文本辨析和语音特性辨析,准确地判断出语音语气,提高添加标点符号的准确性。

[0056] 如图 2 所示,是本发明实施例中训练分类器的流程图,该分类器的训练过程,采用

同时训练文本分类器和声学分类器的方式,充分保证分类的准确性。

[0057] 该流程包括以下步骤:

[0058] 步骤 201,收集文本语料。

[0059] 在具体应用中,可以由人工从网络或书籍上收集各种文本语料,也可以由应用软件从网络上收集各种文本语料,每条语料不仅包含文字内容,而且还包含相应的标点符号。所述文本语料主要以逗号、感叹号、问号和句号为标志,划分为句子单元。所述文本语料包括不同的语气句子类型,如陈述句、感叹句、疑问句等,各句子的长度及声调可以相同或不同。

[0060] 步骤 202,从所述文本语料中提取关键词。

[0061] 具体地,可以针对句子标点符号为句号、问号、感叹号的语句,分别统计其中出现频率较高、而在其他标点对应的文本中出现较少的词语,例如疑问语气中的“为什么”、“吗”、“谁”、“哪个”等,统计出现频率较高的若干个词,构成一维向量空间。

[0062] 其中关键词的挑选准则为:对于三类标点符号 C_1 (句号)、 C_2 (问号)、 C_3 (叹号) 对应的语料,挑选其中每类语料中出现的后验概率最大的若干前 N 个词,公式如下:

$$[0063] \quad Q_i = \arg[\max_{w:topN} p(C_i | w)] = \arg[\max_{w:topN} \frac{p(w | C_i) p(C_i)}{\sum_{i=1}^3 p(w | C_i) p(C_i)}] \quad (1)$$

[0064] W 表示文本语料的个数, C_i 表示不同语气的语料的个数。

[0065] 其中, $i = 1, 2, 3$, $p(C_i | w)$ 为一句语料中找到的关键词属于第 i 类的概率。对于 N 的选取可以根据语料的多少以及识别率的好坏进行不同选择,比如, N 取 20。

[0066] 步骤 203,根据提取的关键词训练生成文本分类器。

[0067] 具体地,可以将提取的关键词以类型为句号、感叹号、问号结尾的文本中出现的关键词组成一个矢量,每种语气仅保留前 N 个最有区分性的词语,例如,从感叹句中挑选的关键词为:啊、吧、唉、哎呀、哦、天哪...;从问句中挑选的关键词为:为什么、谁、哪里、哪一个、什么、怎么样、多少、星期几...;从陈述句中挑选的关键词为:知道了、是的吧、很好的...,共组成长度为 M 的矢量 (M, N 为正整数)。

[0068] 在前面得到的 M 个元素的基础上,对于测试训练集中的所有文本语料,分析每个句子中各个关键词是否出现,如果出现,则在上述 M 维矢量中对应的位置填上 1,其余的位置设为 0。然后利用以这些矢量和对应的标点符号为类别,以最大熵为准则,训练出文本分类器。

[0069] 步骤 204,获取不同的录音人对所述文本语料录音生成的语音文件。

[0070] 比如,对收集的文本语料,选取 5 名男性和 5 名女性录音人,每个人按照日常说话的方式读出所述文本语料的内容,并且根据不同的句型采用不同的语气朗读。同时,保证录音人对于每种语气需要读大体相当数量的文本。对录音后的文件按照 16KHz 采样率、16bit 量化进行 A/D 变换,以 WAV 格式保存该文件。

[0071] 步骤 205,确定所述语音文件中的发音语气。

[0072] 具体地,可以通过对所述语音文件进行语气分析,确定所述语音文件中的发音语气。所述语气分析包括:包括韵律分析和频谱分析。

[0073] 上述韵律分析包括基频提取和规整、一阶、二阶基频差分。其中,基频提取和规整

可以采用 Praat 基频提取算法,计算每一帧位置对应的基频值以及计算该句的基频平均值(基频值为 0 的位置不参与计算),然后将所有基频值减去整体的均值。

[0074] 上述频谱分析包括计算平均频率幅度、共振峰参数提取。其中,计算平均频率幅度即对于每一帧数据,采用离散傅里叶变换(DFT),计算出每一帧数据平均幅度,即公式:

$$[0075] \quad F = \frac{\sum_{i=1}^N A_i F_i}{\sum_{i=1}^N A_i} \quad (2)$$

[0076] 其中, F_i 为离散傅里叶变换时各个谐波频率, A_i 为与各个谐波对应的幅度。提取第一共振峰参数,对于每帧数据,采用线性预测法(LPC)求出 N 阶预测系数,然后计算第一共振峰的频率,最后计算频率幅度和共振峰的一阶差分和二阶差分。

[0077] 对于上述韵律和频谱参数,可以按照时长平分 5 份,对于每一份韵律和频谱参数,对上述每种特征(即前面提到的韵律分析中涉及的参数:一阶二阶基频差分、以及频谱分析中涉及的参数:平均频率幅度和共振峰参数)计算特征的平均值,形成一个 30 维的观察矢量。

[0078] 对于每份录音数据的文本,采用上述步骤 203 生成的文本分类器,计算出该句文本属于各类语气的条件概率,和前面的频谱、韵律参数拼接在一起形成一个 33 维分类特征。

[0079] 步骤 206,以所述发音语气为分类目标训练生成声学分类器。

[0080] 对于每一条语句,对于除静寂段以外每一个音节计算平均频谱幅度、基频,加上各维特征的一阶、二阶差分,组成一个矢量。以各语音文件中的发音语气(包括陈述句、疑问句、感叹句)为分类目标训练隐支持向量机(SVM)分类器,作为声学分类器。

[0081] 需要说明的是,上述步骤 202 至步骤 203 表示的是训练文本分类器的过程,步骤 204 至步骤 206 表示的是声学分类器的过程。在实际应用中,可以根据需要选用,以生成文本分类器或声学分类器。当然,也可以同时生成文本分类器和声学分类器。

[0082] 通过上述过程,可以完成分类器的训练。利用已训练的分类器,可以检测语句的语音及其文本,从而确定该语句的语气类型。

[0083] 如图 3 所示,是本发明实施例中根据已训练的分类器确定语句的语气类型的一种流程图,包括以下步骤:

[0084] 步骤 301,对当前需要识别的语句提取关键词。

[0085] 步骤 302,利用文本分类器计算上述语句属于每个类别的条件概率,同时根据声学分类器对上述语句进行语气分析,包括韵律分析和频谱分析。

[0086] 步骤 303,根据计算得到的条件概率及语气分析结果,确定当前语句的语气类型。

[0087] 相应地,本发明实施例还提供一种实现语音识别中自动添加标点符号的系统,如图 4 所示,是该系统的一种结构示意图。

[0088] 在该实施例中,所述系统包括:

[0089] 信号采集单元 401,用于采集用户语音信号;

[0090] 语音识别单元 402,用于对所述用户语音信号进行语音识别,生成包含多个语句的

文字序列；

[0091] 计算单元 403,用于依次计算所述文字序列中语句间的停顿位置的时长；

[0092] 判断单元 404,用于判断所述时长是否小于预设的门限值；

[0093] 第一标点符号添加单元 405,用于在所述判断单元 404 判断所述时长小于预设的门限值,在所述停顿位置添加逗号；

[0094] 语句类型确定单元 406,用于在所述判断单元 404 判断所述时长大于或等于所述门限值,利用预先生成的分类器,确定所述停顿位置前的语句的语气类型；

[0095] 第二标点符号添加单元 407,用于根据所述语句类型确定单元 406 确定的语句类型在所述停顿位置添加标点符号。

[0096] 上述门限值可以根据不同的语速来确定,比如在正常的语速下设为 0.2s,如果语速较慢,则设为 0.3s,对此本发明实施例不做限定。

[0097] 在本发明实施例中,语句类型可以包括:陈述句、疑问句和感叹句。相应地,上述第二标点符号添加单元 407 具体可以在所述语句类型确定单元 406 确定的语句类型为陈述句时,在所述停顿位置添加句号;为疑问句时,在所述停顿位置添加问号;为感叹句时,在所述停顿位置添加感叹号。

[0098] 本发明实施例实现语音识别中自动添加标点符号的系统,通过对采集的用户语音信号进行语音识别,生成包含多个语句的文字序列;依次计算该文字序列中语句间的停顿位置的时长;如果该时长小于预设的门限值,则在该停顿位置添加逗号;如果该时长大于或等于所述门限值,则确定该停顿位置前的语句的语气类型,并根据确定的语句类型在该停顿位置添加标点符号。从而简单高效地实现标点符号的自动添加,并保证了标点符号添加的准确性和灵活性。

[0099] 在实际应用中,上述语句类型确定单元 406 可以采用多种方式确定所述停顿位置前的语句的语气类型,比如,通过预先训练生成的分类器确定所述停顿位置前的语句的语气类型。该分类器可以由本发明实施例的系统生成,也可以由其他系统生成,对此本发明实施例不做限定。

[0100] 如图 5 所示,是本发明实施例实现语音识别中自动添加标点符号的系统的另一种结构示意图。

[0101] 与图 4 所示实施例不同的是,在该实施例中,所述实现语音识别中自动添加标点符号的系统还包括:语料收集单元 501 和分类器生成单元 502,其中:

[0102] 语料收集单元 501 用于收集文本语料;

[0103] 分类器生成单元 502 用于根据所述文本语料通过训练方式生成分类器,所述分类器包括:文本分类器和/或声学分类器。

[0104] 相应地,在该实施例中,语句类型确定单元 406 利用所述分类器生成单元 502 预先生成的分类器,确定所述停顿位置前的语句的语气类型。

[0105] 需要说明的是,在实际应用中,语料收集单元 501 可以自动从网络上收集各种文本语料,每条语料不仅包含文字内容,而且还包含相应的标点符号。所述文本语料主要以逗号、感叹号、问号和句号为标志,划分为句子单元。所述文本语料包括不同的语气句子类型,如陈述句、感叹句、疑问句等,各句子的长度及声调可以相同或不同。当然,上述语料收集单元 501 是可选的,比如,也可以由人工从网络或书籍上收集所需的各种文本语料并提供给

本发明实施例的系统。

[0106] 在本发明实施例中,上述分类器生成单元 502 可以包括:文本分类器训练单元和/或声学分类器训练单元(未图示),其中:所述文本分类器训练单元,用于根据所述文本语料通过训练方式生成文本分类器;所述声学分类器训练单元,用于根据所述文本语料通过训练方式生成声学分类器。

[0107] 在具体实现中,所述文本分类器训练单元可以包括:关键词提取子单元和第一训练子单元,其中:

[0108] 上述关键词提取子单元用于从上述语料收集单元 501 收集的文本语料中提取关键词;

[0109] 上述第一训练子单元用于根据所述关键词提取子单元 501 提取的关键词训练生成文本分类器。

[0110] 所述声学分类器训练单元可以包括:语音文件获取子单元、语气确定子单元和第二训练子单元,其中:

[0111] 上述语音文件获取子单元,用于获取不同的录音人对所述语料收集单元收集的文本语料录音生成的语音文件;

[0112] 上述语气确定子单元,用于确定所述语音文件中的发音语气;

[0113] 上述第二训练子单元,用于以所述语气确定子单元确定的发音语气为分类目标训练生成声学分类器。

[0114] 利用上述分类器生成单元 502 生成文本分类器和/或声学分类器的详细过程可参照前面本发明实施例实现语音识别中自动添加标点符号的方法中的描述,在此不再赘述。

[0115] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元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。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0116] 以上对本发明实施例进行了详细介绍,本文中应用了具体实施方式对本发明进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及设备;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

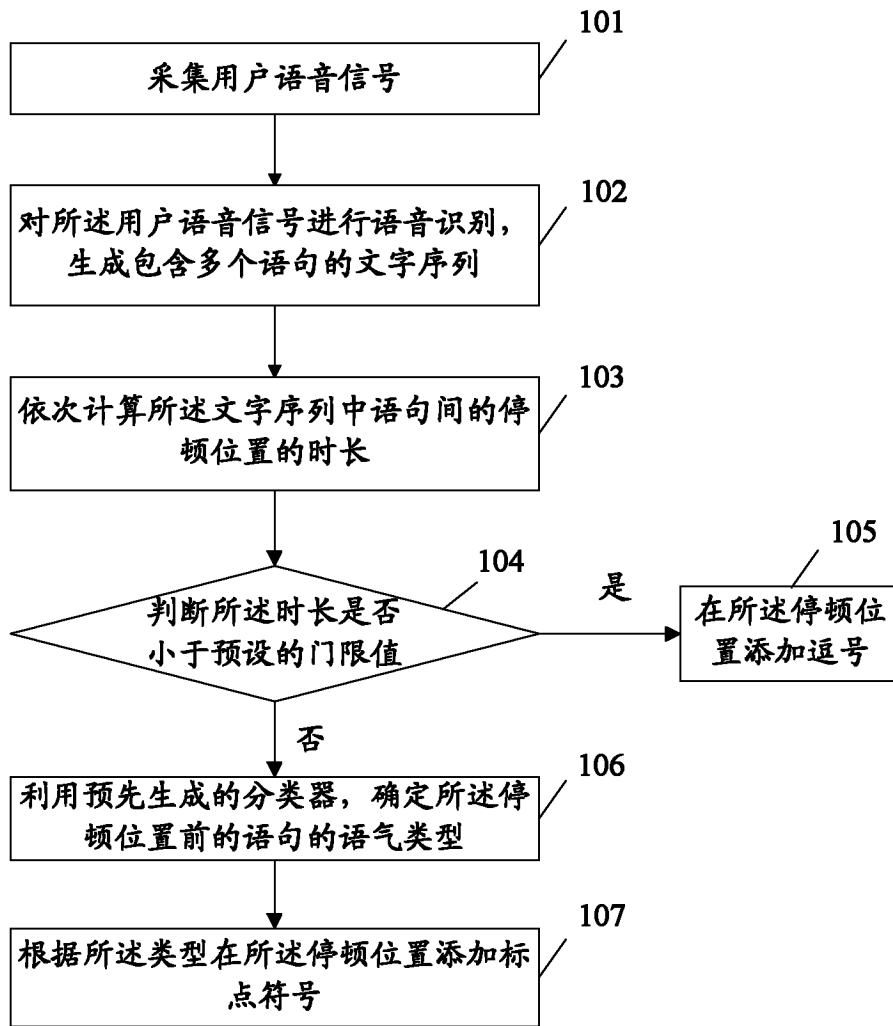


图 1

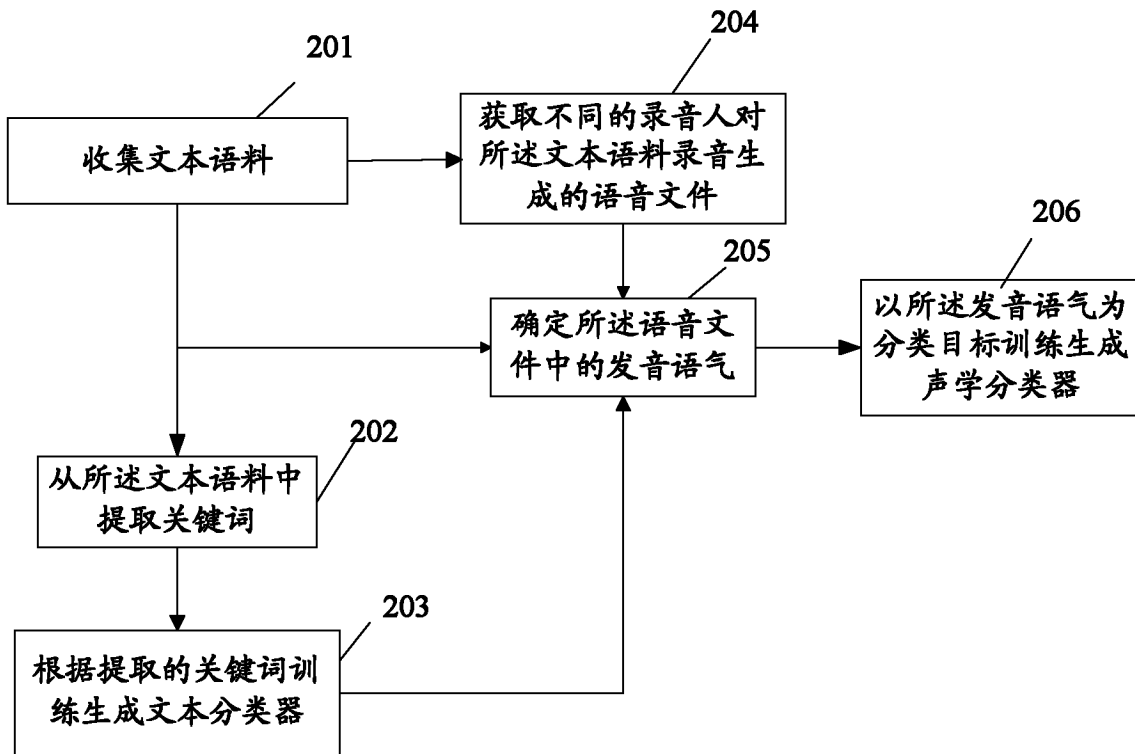


图 2

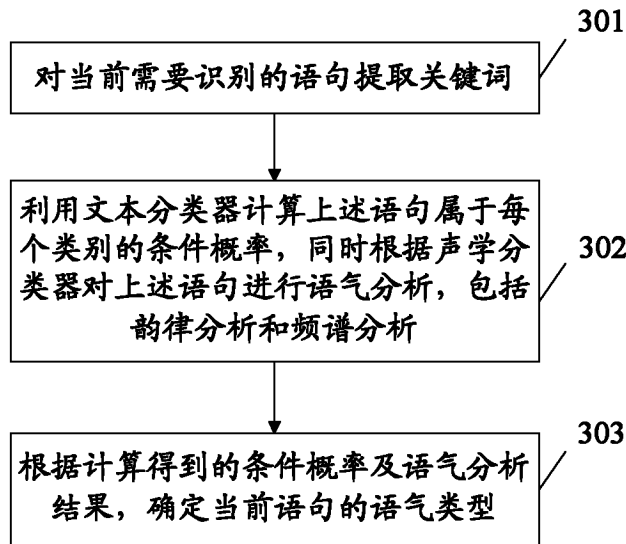


图 3

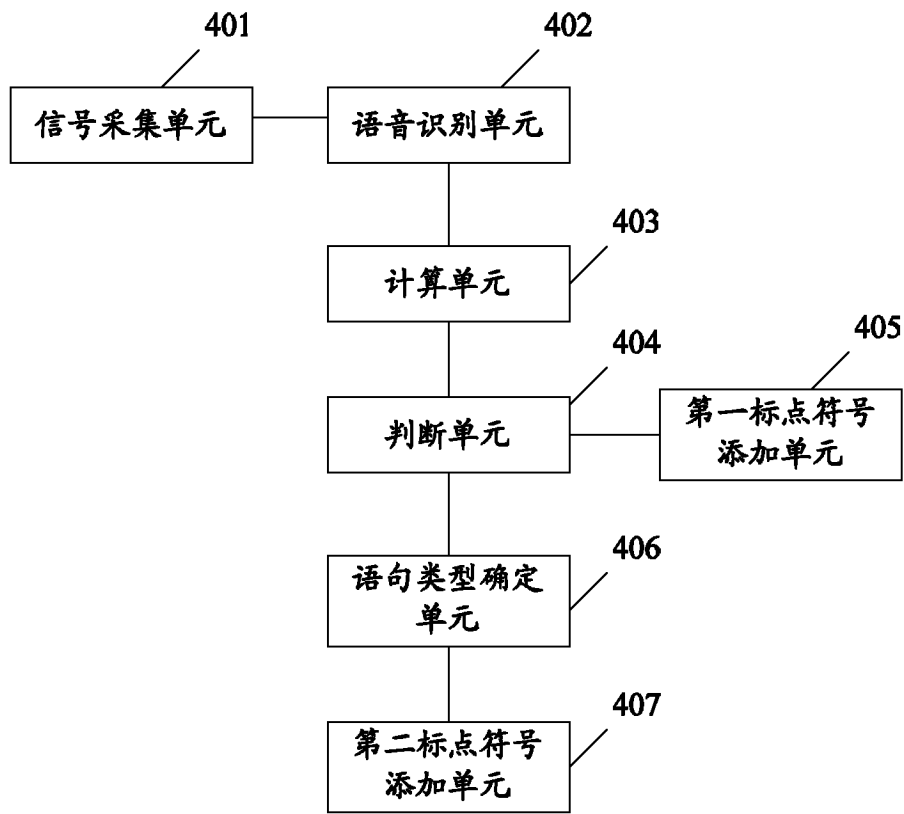


图 4

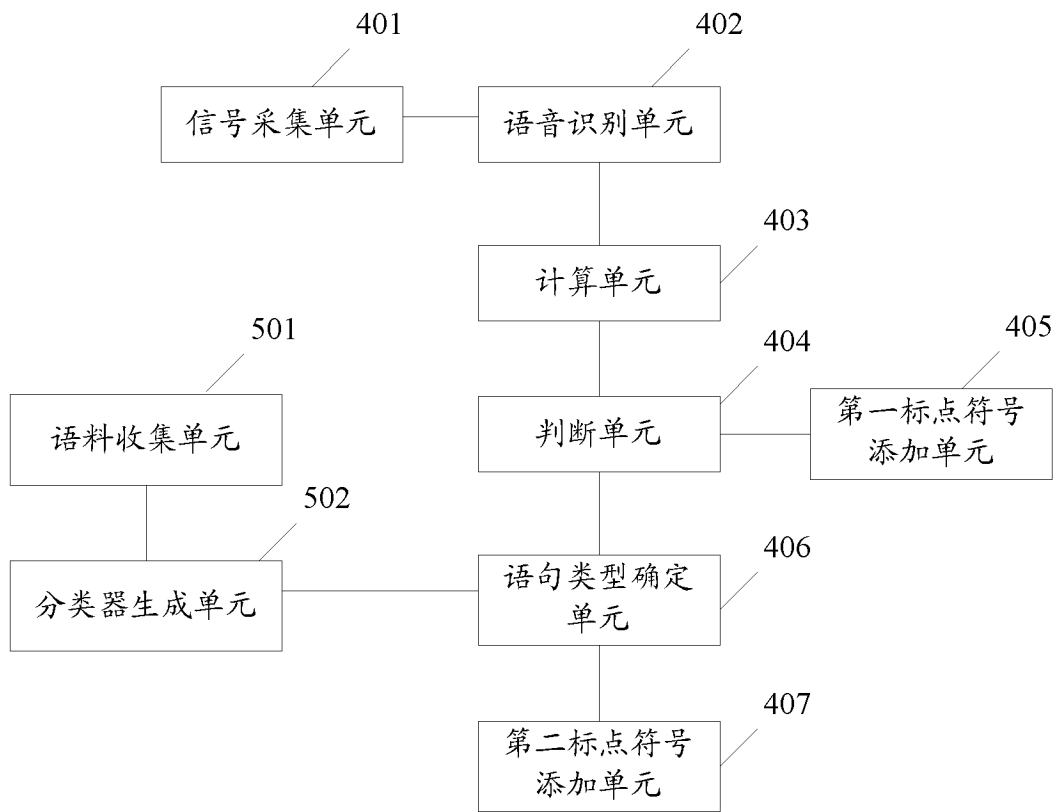


图 5