



(12) 发明专利申请

(10) 申请公布号 CN 102231276 A

(43) 申请公布日 2011. 11. 02

(21) 申请号 201110167390. 8

(22) 申请日 2011. 06. 21

(71) 申请人 北京捷通华声语音技术有限公司

地址 100193 北京市海淀区东北旺西路 8 号
中关村软件园 10 号楼二层 206-1

(72) 发明人 王愈 李健

(74) 专利代理机构 北京润泽恒知识产权代理有
限公司 11319

代理人 苏培华

(51) Int. Cl.

G10L 13/08 (2006. 01)

权利要求书 3 页 说明书 12 页 附图 3 页

(54) 发明名称

一种语音合成单元时长的预测方法及装置

(57) 摘要

本发明提供了一种语音合成单元时长的预测方法和装置，包括：针对上下文环境参数，采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得初始时长预测结果；采用决策树-高斯混合模型对所述初始时长预测结果进行分配，得到分配后的时长预测结果。本发明能够提高时长预测结果的准确性，使得从语音合成系统中合成出的语音具备真实的韵律感。

针对上下文环境参数，采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得初始时长预测结果

采用决策树-
高斯混合模型对所述初始时长预测结果进行分
配，得到分配后的时长预测结果

301

302

1. 一种逐步线性回归的时长预测模型的训练方法, 其特征在于, 包括 :

建立初始的线性回归的时长预测模型;

在迭代所述线性回归的时长预测模型的过程中, 通过评价每轮的时长预测模型选择上下文环境参数, 最终得到最优时长预测模型。

2. 根据权利要求 1 所述的方法, 其特征在于, 所述在迭代所述线性回归的时长预测模型的过程中, 通过评价每轮的时长预测模型选择上下文环境参数, 最终得到最优时长预测模型的步骤, 包括 :

步骤 1 : 选中常参数, 并将其加入已选参数集;

步骤 2 : 进行迭代, 其中, 在每轮迭代的过程中, 在已选参数的基础上选出对进一步提升预测准确度作用最大的未选参数, 并加入已选参数集;

步骤 3 : 利用新的已选参数集, 获得当前轮逐步线性回归的时长预测模型;

步骤 4 : 判断当前轮逐步线性回归的时长预测模型是否最优, 若是, 则以当前逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型, 否则, 返回执行步骤 2。

3. 根据权利要求 2 所述的方法, 其特征在于, 所述判断当前轮逐步线性回归的时长预测模型是否最优的步骤, 包括 :

若当前轮逐步线性回归的时长预测模型相对于上一轮逐步线性回归的时长预测模型, 二者预测误差样本方差的差小于等于特定阈值, 则以当前轮逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型;

若二者预测误差样本方差的差大于特定阈值, 则返回执行步骤 2。

4. 根据权利要求 3 所述的方法, 其特征在于, 所述线性回归的时长预测模型的表达式如下 :

$$\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 \end{cases}$$

其中,

X 为上下文环境参数矩阵, X 的列数为上下文环境参数的数目, 行数为语音合成单元的样本数目, X 具体可以表述为 :

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

Y 为 X 的时长预测矩阵, Y 具体可以表述为 :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

β 为回归模型的回归系数, 具体可以表述为 :

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

ϵ 为预测误差, 具体可以表述为 :

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

σ 为 ε 方差的无偏估计 :

$$\sigma^2 = MSE = \frac{1}{n-2} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2 = Var(\varepsilon)$$

5. 根据权利要求 1 至 4 中任一项所述的方法, 其特征在于 :

在迭代所述线性回归的时长预测模型的过程中, 时长预测模型的估计和评价采用不同的样本集。

6. 一种语音合成单元时长的预测方法, 其特征在于, 包括 :

针对上下文环境参数, 采用逐步线性回归的时长预测模型, 对语音合成单元的时长进行初始预测, 获得初始时长预测结果;

采用决策树 - 高斯混合模型对所述初始时长预测结果进行分配, 得到分配后的时长预测结果。

7. 根据权利要求 6 所述的方法, 其特征在于, 所述采用决策树 - 高斯混合模型对所述初始时长预测结果进行分配的步骤, 包括 :

针对上下文环境参数, 采用决策树 - 高斯混合模型, 对语音合成单元及各语音合成单元子状态的时长进行预测, 获得语音合成单元时长和语音合成单元各子状态时长的缩放比例;

根据语音合成单元时长和语音合成单元各子状态时长的缩放比例, 将所述初始时长预测结果进行等比例缩放, 获得语音合成单元各子状态的时长预测结果。

8. 一种逐步线性回归的时长预测模型的训练装置, 其特征在于, 包括 :

建立模块, 用于建立初始的线性回归的时长预测模型; 及

优化模块, 用于在迭代所述线性回归的时长预测模型的过程中, 通过评价每轮的时长预测模型选择上下文环境参数, 最终得到最优时长预测模型。

9. 一种语音合成单元时长的预测装置, 其特征在于, 包括 :

初始时长预测模块, 用于针对上下文环境参数, 采用逐步线性回归的时长预测模型, 对语音合成单元的时长进行初始预测, 获得初始时长预测结果;

分配模块, 用于采用决策树 - 高斯混合模型对所述初始时长预测结果进行分配, 得到分配后的时长预测结果。

10. 根据权利要求 9 所述的装置, 其特征在于, 所述分配模块, 包括 :

子状态预测单元, 用于针对上下文环境参数, 采用决策树 - 高斯混合模型, 对语音合成单元及各语音合成单元子状态的时长进行预测, 获得语音合成单元时长和语音合成单元各

子状态时长的缩放比例；

缩放单元，用于根据语音合成单元时长和语音合成单元各子状态时长的缩放比例，将所述初始时长预测结果进行等比例缩放，获得语音合成单元各子状态的时长预测结果。

一种语音合成单元时长的预测方法及装置

技术领域

[0001] 本发明涉及信息处理技术领域,特别是涉及一种逐步线性回归的时长预测模型的训练方法及装置、一种语音合成单元时长的预测方法及装置。

背景技术

[0002] 在语音合成系统 (Text-to-Speech, TTS) 中,语音合成单元时长的预测生成是必不可少的步骤,对合成语音的韵律听感有着至关重要的作用。

[0003] 根据语音学与音系学理论,语音合成单元的时长等特性决定于其所处的上下文环境。对语音时长的预测,本质上是从上下文环境参数的取值空间到时长取值空间的映射。对此种映射关系的分析建模方法,现有的时长预测方法通常采用决策树 - 高斯混合模型,确定与之最接近的近似映射。

[0004] 但是,现有的时长预测方法存在一个显著的缺点 :采用决策树 - 高斯混合模型来预测时长,所述预测首先对上下文环境参数的取值空间进行粗分类,然后用单一的均值来刻画各子类空间,在这两个过程中都存在着过平均化。

[0005] 下面以一个实例做说明 :比如“们”字,在“我们”中和在“我们的”中两种情况下,相应的上下文环境都属于“词中”,只是在词中的位置不同。在基于决策树建立的决策树 - 高斯混合模型中,基于决策树的聚类因为受到树节点数目的限制,只能选择最显著的分类标准进行粗分类,有可能将这两种情况同归为“词中”这一类,从而抹煞了二者各自的个性;在此类别内,使用决策树 - 高斯混合模型建模,是用单一的均值来刻画整个子类,进一步抹煞了各样本具体的个性。

[0006] 总之,需要本领域技术人员迫切解决的一个技术问题就是 :如何提供一种时长预测模型的训练方法,以提高时长预测结果的准确性。

发明内容

[0007] 本发明所要解决的技术问题是提供一种逐步线性回归的时长预测模型的训练方法及装置、一种语音合成单元时长的预测方法及装置,能够提高时长预测结果的准确性,使得从语音合成系统中合成出的语音具备真实的韵律感。

[0008] 为了解决上述问题,本发明公开了一种逐步线性回归的时长预测模型的训练方法,包括 :

[0009] 建立初始的线性回归的时长预测模型;

[0010] 在迭代所述线性回归的时长预测模型的过程中,通过评价每轮的时长预测模型选择上下文环境参数,最终得到最优时长预测模型。

[0011] 优选的,所述在迭代所述线性回归的时长预测模型的过程中,通过评价每轮的时长预测模型选择上下文环境参数,最终得到最优时长预测模型的步骤,包括 :

[0012] 步骤 1 :选中常参数,并将其加入已选参数集;

[0013] 步骤 2 :进行迭代,其中,在每轮迭代的过程中,在已选参数的基础上选出对进一

步提升预测准确度作用最大的未选参数，并加入已选参数集；

[0014] 步骤 3：利用新的已选参数集，获得当前轮逐步线性回归的时长预测模型；

[0015] 步骤 4：判断当前轮逐步线性回归的时长预测模型是否最优，若是，则以当前逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型，否则，返回执行步骤 2。

[0016] 优选的，所述判断当前轮逐步线性回归的时长预测模型是否最优的步骤，包括：

[0017] 若当前轮逐步线性回归的时长预测模型相对于上一轮逐步线性回归的时长预测模型，二者预测误差样本方差的差小于等于特定阈值，则以当前轮逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型；

[0018] 若二者预测误差样本方差的差大于特定阈值，则返回执行步骤 2。

[0019] 优选的，所述线性回归的时长预测模型的表达式如下：

$$\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 \end{cases}$$

[0021] 其中，

[0022] X 为上下文环境参数矩阵， X 的列数为上下文环境参数的数目，行数为语音合成单元的样本数目， X 具体可以表述为：

[0023]

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

[0024] Y 为 X 的时长预测矩阵， Y 具体可以表述为：

$$[0025] Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

[0026] β 为回归模型的回归系数，具体可以表述为：

$$[0027] \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

[0028] ε 为预测误差，具体可以表述为：

$$[0029] \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

[0030] σ 为 ε 方差的无偏估计：

$$[0031] \sigma^2 = MSE = \frac{1}{n-2} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2 = Var(\varepsilon)$$

[0032] 优选的，在迭代所述线性回归的时长预测模型的过程中，时长预测模型的估计和

评价采用不同的样本集。

[0033] 另一方面，本发明还公开了一种语音合成单元时长的预测方法，包括：

[0034] 针对上下文环境参数，采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得初始时长预测结果；

[0035] 采用决策树－高斯混合模型对所述初始时长预测结果进行分配，得到分配后的时长预测结果。

[0036] 优选的，所述采用决策树－高斯混合模型对所述初始时长预测结果进行分配的步骤，包括：

[0037] 针对上下文环境参数，采用决策树－高斯混合模型，对语音合成单元及各语音合成单元子状态的时长进行预测，获得语音合成单元时长和语音合成单元各子状态时长的缩放比例；

[0038] 根据语音合成单元时长和语音合成单元各子状态时长的缩放比例，将所述初始时长预测结果进行等比例缩放，获得语音合成单元各子状态的时长预测结果。

[0039] 另一方面，本发明还公开了一种逐步线性回归的时长预测模型的训练装置，包括：

[0040] 建立模块，用于建立初始的线性回归的时长预测模型；及

[0041] 优化模块，用于在迭代所述线性回归的时长预测模型的过程中，通过评价每轮的时长预测模型选择上下文环境参数，最终得到最优时长预测模型。

[0042] 另一方面，本发明还公开了一种语音合成单元时长的预测装置，包括：

[0043] 初始时长预测模块，用于针对上下文环境参数，采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得初始时长预测结果；

[0044] 分配模块，用于采用决策树－高斯混合模型对所述初始时长预测结果进行分配，得到分配后的时长预测结果。

[0045] 优选的，所述分配模块，包括：

[0046] 子状态预测单元，用于针对上下文环境参数，采用决策树－高斯混合模型，对语音合成单元及各语音合成单元子状态的时长进行预测，获得语音合成单元时长和语音合成单元各子状态时长的缩放比例；

[0047] 缩放单元，用于根据语音合成单元时长和语音合成单元各子状态时长的缩放比例，将所述初始时长预测结果进行等比例缩放，获得语音合成单元各子状态的时长预测结果。

[0048] 与现有技术相比，本发明具有以下优点：

[0049] 本发明提供一种逐步线性回归的时长预测模型，由于对语音时长的预测，本质上是从上下文环境参数的取值空间到时长取值空间的映射，而回归预测能够直接描述这种映射关系，而逐步线性回归“逐步参数优选”的策略，旨在兼顾精简的同时逐步逼近真实映射关系；因此，所述逐步线性回归的时长预测模型能够最大程度地逼近从 X(上下文环境参数的取值空间) 到 Y(时长取值空间) 的映射，相对于现有的决策树－高斯混合模型，所述逐步线性回归的时长预测模型具有更加准确的时长预测能力。

[0050] 其次，语音参数的生成是以语音合成单元的子状态为单位进行的，其先决条件之一是语音合成单元各子状态的时长，而所述逐步线性回归的时长预测模型生成的时长值只

是具体到语音合成单元这一级别，并没有细分到其子状态层级；因此，本发明在进行语音合成单元时长的预测时，首先采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得初始时长预测结果，然后采用决策树－高斯混合模型对所述初始时长预测结果进行分配，获得语音合成单元各子状态的时长预测结果；所述逐步线性回归的时长预测模型所具有的准确的时长预测能力，能够保证所述初始时长预测结果和语音合成单元各子状态的时长预测结果的准确性。

[0051] 再者，在所述逐步线性回归的时长预测模型的训练过程中，模型的估计和评价可以使用两组不同的样本集，辅之以参数集合的精简，可以有效地减小模型对训练数据的过度拟合，从而提高预测模型的可外推性。

附图说明

[0052] 图 1 是本发明一种逐步线性回归的时长预测模型的训练方法实施例的流程图；

[0053] 图 2 是本发明一种时长预测模型的训练方法中迭代算法的流程图；

[0054] 图 3 是本发明一种语音合成单元时长的预测方法实施例的流程图；

[0055] 图 4 是本发明通过逐步线性回归模型预测出的时长值进行语音合成的流程图；

[0056] 图 5 是本发明一种逐步线性回归的时长预测模型的训练装置实施例的结构图；

[0057] 图 6 是本发明一种语音合成单元时长的预测装置实施例的结构图。

具体实施方式

[0058] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0059] 现有的时长预测模型的训练方法，采用决策树－高斯混合模型预测时长，不能获得准确的时长预测结果的原因在于，决策树－高斯混合模型是在决策树的基础上建立起来的。由于基于决策树的聚类受到树节点数目的限制，只能选择最显著的分类标准进行粗分类；这将使得通过决策树－高斯混合模型来预测时长，是用单一时长的均值来刻画整个子类时长值，从而抹煞了某一类别中各个样本具体个性之间的差异；这样得到的时长预测结果不准确，且过于平均化。

[0060] 本专利发明人注意了这一点，因此创造性地提出了本发明实施例的核心构思之一，也即，采用逐步线性回归的时长预测模型进行语音时长的预测；由于对语音时长的预测，本质上是从上下文环境参数的取值空间到时长取值空间的映射，而回归预测能够直接描述这种映射关系，而逐步线性回归“逐步参数优选”的策略，旨在兼顾精简的同时逐步逼近真实映射关系。

[0061] 参照图 1，示出了本发明一种逐步线性回归的时长预测模型的训练方法实施例的流程图，具体可以包括：

[0062] 步骤 101、建立初始的线性回归的时长预测模型；

[0063] 本发明实施例中，所述逐步线性回归的时长预测模型是通过回归分析方法建立起来的时长预测模型；所述逐步线性回归的时长预测模型是对从上下文环境参数的取值空间到时长取值空间的映射关系最直观的分析建模方法，确定与之最接近的近似映射。

[0064] 在本发明的逐步线性回归的时长预测模型中 (Duration Prediction with Stepwise Linear Regression), 对于每类语音合成单元, 可以使用如下公式定义的多元线性回归模型预测相应的时长 :

[0065] 如果假定上下文环境参数与时长之间的关系为线性关系, 则二者的映射关系可以表示为 :

$$[0066] Y = X \beta \quad (1)$$

[0067] 其中, X 为上下文环境参数矩阵, X 的列数为上下文环境参数的数目, 行数为语音合成单元的样本数目, X 具体可以表述为 :

[0068]

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad (2)$$

[0069] Y 为 X 的时长预测矩阵, Y 具体可以表述为 :

$$[0070] Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (3)$$

[0071] β 为回归模型的回归系数, 具体可以表述为 :

$$[0072] \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad (4)$$

[0073] 通常, 上下文环境参数的数目为几百, 而语音合成单元的样本数目数以万计, 在行数多于列数的情况下, 此方程无精确解, 只能寻找最优近似解。也就是

$$[0074] Y = X \beta + \varepsilon \quad (5)$$

[0075] ε 是预测误差, 寻找最优的 $\hat{\beta}$ 使得整体预测误差最小, $\hat{\beta}$ 为 β 的预测。这一方法就是回归预测。从几何意义上讲, 就是寻找一条直线, 能够对全体样本点做最佳拟合。在求解 β 时, 并不需要 ε 参与, ε 只作为事后的统计评价参数 :

$$[0076] \varepsilon = Y - X \hat{\beta} \quad (6)$$

[0077] 可以证明, ε 的均值为零 $E(\varepsilon) = \bar{\varepsilon} = 0$, 进而可得预测误差的样本方差 :

$$[0078] MSE = \frac{1}{n-2} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2 = Var(\varepsilon) = \sigma^2 \quad (7)$$

[0079] 可以证明 MSE 是 ε 方差的无偏估计

$$[0080] \sigma^2 = Var(\varepsilon) = MSE \quad (8)$$

[0081] σ^2 综合体现了预测误差的总和大小及变异程度。一个好的回归拟合方程, 其预测误差总和应越小越好 : 预测误差越小, 拟合值与观测值越接近, 各观测点在拟合直线周围聚集的紧密程度越高, 也就是说, 回归模型对 y 的解释能力越强 ; 另一方面, σ^2 越小, 预测误

差值的变异程度越小。由于预测误差的样本均值为零,所以其离散范围越小,拟合的模型就越精确。总之,使用 σ^2 作为回归模型的评价标准,是简捷有效的。

[0082] 因此,本发明建立逐步线性回归的时长预测模型:

$$\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 \end{cases} \quad (9)$$

[0084] 步骤 102、在迭代所述线性回归的时长预测模型的过程中,通过评价每轮的时长预测模型选择上下文环境参数,最终得到最优时长预测模型。

[0085] 总体上讲,引入的上下文环境参数 (X 的列数) 越多,逐步线性回归的时长预测模型的刻画能力越强越细腻,但也并非多多益善:首先,过多的参数会引入冗余,造成巨大的不必要的计算代价;其次,部分参数之间存在相关性,可能是正面的也可能是负面的,从而造成 $1+1 < 2$ 的结果;此外,由于回归分析必须在 X 的列数(远) 小于行数的前提下进行,过多的参数 (X 的列数) 意味着需要更多的训练样本 (X 的行数),而语料库对语音合成单元上下文环境的取值空间覆盖不足、不均衡,是无法避免的问题,从而导致 X 的行数不足,Y 取值不均衡,继而导致模型过拟合(过度贴近于训练数据,而对训练集之外的数据缺乏描述力)与偏倚。总之,如果能够优选出对时长预测的贡献度较大的上下文环境参数,就可以兼顾准确度、效率以及可外推性 (extrapolation)。

[0086] 究竟哪些上下文环境参数对时长预测起主导作用?已有的一种方法通常基于专家知识主观指定,显然这种方法过于主观和片面。为此,已有的另一种方法使用有效的统计学指标来分别评价各参数的重要性,乃至参数间两两的交互作用,然后基于评价结果主观选择最重要的一些参数。这种方法的局限性在于它是静态的:在统一的前提条件下单独评价每个参数,即使是两两交互也出一辙。此外,参数间的相关性也并非两两交互这样简单。

[0087] 针对已有方法的局限性,本发明提供了一种步步为营、逐步逼近的动态过程,具体而言,每轮迭代,都选择当前可选的最重要的参数,而评价所谓重要的标准是在已选参数集合的基础上加入该参数后,预测误差的 σ^2 最小。这种情况下,由于综合考虑了待入选参数与全部已选参数之间的多角交互作用,且每轮优选都是在上一步达到最优状态的前提下进行,故每轮迭代获得的最小的 σ^2 能够体现该轮可达的最优状态,只要新一轮的 σ^2 比上一轮的 σ^2 有明显下降,就表示新一轮迭代有价值,进一步迭代下去还有所可为;反之,如果新一轮的 σ^2 比上一轮的 σ^2 下降不明显,甚至不降反升,则表明已经进入冗余状态,继续迭代下去徒劳无益,甚至适得其反。

[0088] 在本发明的一种优选实施例中,可以从一个初始常参数开始,逐步引入待选参数中对 Y 作用最显著的参数;重复这个过程,直至剩余方差无下降或下降不明显为止。

[0089] 相应地,所述步骤 102 可以进一步包括:

[0090] 步骤 1:选中常参数,并将其加入已选参数集;

[0091] 步骤 2:进行迭代,其中,在每轮迭代的过程中,在已选参数的基础上选出对进一步提升预测准确度作用最大的未选参数,并加入已选参数集;

[0092] 步骤 3:利用新的已选参数集,获得当前轮逐步线性回归的时长预测模型;

[0093] 步骤 4:判断当前轮逐步线性回归的时长预测模型是否最优,若是,则以当前逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型,否则,返回执行步骤 2。

[0094] 在本发明的一种优选实施例中,在迭代所述线性回归的时长预测模型的过程中,

时长预测模型的估计和评价可以采用不同的样本 $\{X_{train}, Y_{train}\}$ 和 $\{X_{evaluate}, Y_{evaluate}\}$ 。

[0095] 其中,在依据 $Y = X\beta + \epsilon$ 进行时长预测模型的估计的过程中,用于求解 β 的 X 和相应的 Y 称为估计样本(训练集),求解得到最优的 $\hat{\beta}$ 后,可以继而统计 ϵ 并最终获得 σ^2 ,以评价 $\hat{\beta}$ 对这组数据集的描述力。并且,在模型评价的过程中使用了另外一组独立的数据集,用训练集之外的数据评价训练得到的模型,辅之以参数集合的精简,可以有效地减小模型对训练数据的过度拟合,从而提高预测模型的可外推性。最终结果是,在较为普适的范畴内,能够优选出较为重要的上下文环境参数。

[0096] 在本发明的另一种优选实施例中,所述判断当前轮逐步线性回归的时长预测模型是否最优的步骤,可以进一步包括:

[0097] 若当前轮逐步线性回归的时长预测模型相对于上一轮逐步线性回归的时长预测模型,二者预测误差样本方差的差小于等于特定阈值,则以当前轮逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型;

[0098] 若二者预测误差样本方差的差大于特定阈值,则返回执行步骤 2。

[0099] 参照图 2,示出了本发明一种时长预测模型的训练方法中迭代算法的流程图,该迭代算法从一个初始常参数开始,逐步引入待选参数中对 Y 作用最显著的参数,重复这个过程,直至剩余方差无下降或下降不明显为止。在此过程中,估计回归参数和评价剩余标准差分别使用两组不同的样本 $\{X_{train}, Y_{train}\}$ 和 $\{X_{evaluate}, Y_{evaluate}\}$;该迭代算法具体可以包括:

[0100] 步骤 201、 $S_{selected}$ 初始化,随之调整 $X_{train, selected}$, $X_{evaluate, selected}$, 计算 $\beta_{selected}$ 和 $\sigma_{selected}$;

[0101] 其中 $S_{candidate}$ 为待选参数的集合,其体现在矩阵 X_{train} 的最大列序号,也即,矩阵 X_{train} 的最大列序号的初始值为全部上下文环境参数的数目; $S_{selected}$ 为已选中参数的集合, $X_{train, selected}$ 和 $X_{evaluate, selected}$ 分别为 X_{train} 和 $X_{evaluate}$ 的子矩阵, $\beta_{selected}$ 是用 $\{X_{train, selected}, Y_{train}\}$ 估计出的回归参数, $\sigma_{selected}$ 为在此情况下用 $\{X_{evaluate, selected}, Y_{evaluate}\}$ 计算出的剩余标准差。

[0102] 步骤 202、对于 $S_{candidate}$ 中的各元素 C_i ,选择最小的 σ_i 所对应的列,加入 $S_{candidate}$ 中,更新 $\sigma_{selected}$ 为最小的 σ_i ,并计算下降值 $\Delta_{selected}$;

[0103] 其中,

[0104] $\Delta_{selected}$ 为 ϵ 的标准差减去 σ ;

[0105] 在 $X_{train, selected}$, $X_{evaluate, selected}$ 中加入此列,计算 β_i 和 σ_i 。

[0106] 步骤 203、通过判断 $\Delta_{selected}$ 是否小于特定阈值,来判断当前的 $\beta_{selected}$ 即为最终的模型参数;

[0107] 在本发明的优选实施例中,所述通过判断 $\Delta_{selected}$ 是否小于等于特定阈值,来判断当前的 $\beta_{selected}$ 即为最终的模型参数的步骤,具体可以包括:

[0108] 子步骤 D1:若 $\Delta_{selected}$ 小于等于特定阈值时,停止迭代,以当前逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型;

[0109] 子步骤 D2:若 $\Delta_{selected}$ 大于特定阈值时,返回重复执行步骤 202;直到新一轮迭代中的 $\Delta_{selected}$ 小于特定阈值时,停止迭代,以当前逐步线性回归的时长预测模型作为逐步线性回归的最优时长预测模型。

[0110] 为使本领域技术人员更好地理解本发明，下面以一个具体的实例来说明上述时长预测模型的训练方法中算法流程的步骤，具体可以包括：

[0111] 子步骤 E1 : $S_{selected}$ 初始化，随之调整 $X_{train, selected}$, $X_{evaluate, selected}$, 计算 $\beta_{selected}$ 和 $\sigma_{selected}$ ；

[0112] 若已知，

$$[0113] Y_{train} = \begin{bmatrix} y_1^t \\ y_2^t \\ y_3^t \\ y_4^t \end{bmatrix}, X_{train} = \begin{bmatrix} 1 & x_{11}^t & x_{12}^t \\ 1 & x_{21}^t & x_{22}^t \\ 1 & x_{31}^t & x_{32}^t \\ 1 & x_{41}^t & x_{42}^t \end{bmatrix} Y_{evaluate} = \begin{bmatrix} y_1^e \\ y_2^e \\ y_3^e \\ y_4^e \end{bmatrix}, X_{evaluate} = \begin{bmatrix} 1 & x_{11}^e & x_{12}^e \\ 1 & x_{21}^e & x_{22}^e \\ 1 & x_{31}^e & x_{32}^e \\ 1 & x_{41}^e & x_{42}^e \end{bmatrix}$$

[0114] 初始化 $S_{selected} = \{0\}$ 为 X 的常数列，

$$[0115] X_{train, selected} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} X_{evaluate, selected} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

[0116] 待选参数集 $S_{candidate} = \{1, 2\}$ ；

[0117] 由 $Y_{train} = X_{train, selected} \beta_{selected}$ 解出最佳的 $\beta_{selected}$ 。

[0118] 由 $\varepsilon = Y_{evaluate} - X_{evaluate, selected} \beta_{selected}$ 计算出 ε 的标准差 $\sigma_{selected}$ 。

[0119] 子步骤 E2、对于 $S_{candidate}$ 中的各元素 C_i , 选择最小的 σ_i 所对应的列, 加入 $S_{candidate}$ 中, 更新 $\sigma_{selected}$ 为最小的 σ_i , 并计算下降值 $\Delta_{selected}$, 迭代开始；

[0120] 对于 $S_{candidate}$ 中的第一个候选列 1, 调整 X 中的相应列

$$[0121] X_{evaluate, selected} = \begin{bmatrix} 1 & x_{11}^e \\ 1 & x_{21}^e \\ 1 & x_{31}^e \\ 1 & x_{41}^e \end{bmatrix} X_{train, selected} = \begin{bmatrix} 1 & x_{11}^t \\ 1 & x_{21}^t \\ 1 & x_{31}^t \\ 1 & x_{41}^t \end{bmatrix}$$

[0122] 由 $Y_{train} = X_{train, selected} \beta^1$ 解出最佳的 β^1 。

[0123] 由 $\varepsilon = Y_{evaluate} - X_{evaluate, selected} \beta^1$ 计算出 ε 的标准差 σ_1 。

[0124] 对于 $S_{candidate}$ 中的第二个候选列 2, 调整 X 中的相应列

$$[0125] X_{evaluate, selected} = \begin{bmatrix} 1 & x_{12}^e \\ 1 & x_{22}^e \\ 1 & x_{32}^e \\ 1 & x_{42}^e \end{bmatrix} X_{train, selected} = \begin{bmatrix} 1 & x_{12}^t \\ 1 & x_{22}^t \\ 1 & x_{32}^t \\ 1 & x_{42}^t \end{bmatrix}$$

[0126] 由 $Y_{train} = X_{train, selected} \beta^2$ 解出最佳的 β^2 。

[0127] 由 $\varepsilon = Y_{evaluate} - X_{evaluate, selected} \beta^2$ 计算出 ε 的标准差 σ_2 。

[0128] 假设 $\sigma_1 > \sigma_2$, 则本轮迭代选中第二列, $S_{selected} = \{0, 2\}$, $S_{candidate} = \{1\}$, 到目前为止的 X 定型为

$$[0129] \quad X_{\text{train},\text{selected}} = \begin{bmatrix} 1 & x_{12}^t \\ 1 & x_{22}^t \\ 1 & x_{32}^t \\ 1 & x_{42}^t \end{bmatrix} \quad X_{\text{evaluate},\text{selected}} = \begin{bmatrix} 1 & x_{12}^e \\ 1 & x_{22}^e \\ 1 & x_{32}^e \\ 1 & x_{42}^e \end{bmatrix}$$

[0130] $\Delta_{\text{selected}} = \sigma_{\text{selected}} - \sigma_2$, $\sigma_{\text{selected}} = \sigma_2$, 假设 Δ_{selected} 还未小于预设的阈值, 则继续下一轮迭代;

[0131] 对于 $S_{\text{candidate}}$ 中唯一的第一个候选列 1, 调整 X 中的相应列

$$[0132] \quad X_{\text{train},\text{selected}} = \begin{bmatrix} 1 & x_{11}^t & x_{12}^t \\ 1 & x_{21}^t & x_{22}^t \\ 1 & x_{31}^t & x_{32}^t \\ 1 & x_{41}^t & x_{42}^t \end{bmatrix} \quad X_{\text{evaluate},\text{selected}} = \begin{bmatrix} 1 & x_{11}^e & x_{12}^e \\ 1 & x_{21}^e & x_{22}^e \\ 1 & x_{31}^e & x_{32}^e \\ 1 & x_{41}^e & x_{42}^e \end{bmatrix}$$

[0133] 由 $Y_{\text{train}} = X_{\text{train},\text{selected}} \beta^1$ 解出最佳的 β^1 ;

[0134] 由 $\varepsilon = Y_{\text{evaluate}} - X_{\text{evaluate},\text{selected}} \beta^1$ 计算出 ε 的标准差 σ_1 。

[0135] 子步骤 E3、通过判断 Δ_{selected} 是否小于特定阈值或者不降, 来判断上述逐步线性回归的时长预测模型是否为逐步线性回归的最优时长预测模型;

[0136] $\Delta_{\text{selected}} = \sigma_{\text{selected}} - \sigma_1$, 若 Δ_{selected} 大于等于零, 则说明加入第一列后的剩余误差大于等于原有的剩余误差, 此时 Δ_{selected} 不降, 说明当前的 β_{selected} 即为最终的模型参数, 因此, 本轮迭代无产出, 终止迭代; 并且, 将当前模型为逐步线性回归的最优时长预测模型。

[0137] 上述不降是一种特例, 表明迭代过程已经进入冗余状态。除了不降外, 本发明还可以通过特定阈值来判断新一轮的 σ^2 相对于上一轮的 σ^2 , 是否下降明显, 具体地, 如果 Δ_{selected} 小于特定阈值, 则说明下降不明显, 也表明迭代过程已经进入冗余状态; 说明当前的 β_{selected} 即为最终的模型参数, 因此, 本轮迭代无产出, 终止迭代; 并且, 将当前模型为逐步线性回归的最优时长预测模型。

[0138] 当然, 本领域技术人员可以根据实际需要, 设置该特定阈值的值, 如 0.001, 0.002 等, 本发明的宗旨是通过判断迭代是否进入冗余状态, 来判断当前逐步线性回归的时长预测模型是否为最优, 而不会对特定阈值的值加以限制。

[0139] 总之, 本发明提供了一种逐步线性回归的时长预测模型的训练方法, 所述逐步线性回归的时长预测模型, 能够最大程度地逼近从 X (上下文环境参数的取值空间) 到 Y (时长取值空间) 的映射, 从而使得能够获得更加准确的时长预测结果。

[0140] 参照图 3, 示出了本发明一种语音合成单元时长的预测方法实施例的流程图, 具体可以包括:

[0141] 步骤 301、针对上下文环境参数, 采用逐步线性回归的时长预测模型, 对语音合成单元的时长进行初始预测, 获得初始时长预测结果;

[0142] 步骤 302、采用决策树 - 高斯混合模型对所述初始时长预测结果进行分配, 得到分配后的时长预测结果。

[0143] 根据语音学与音系学理论, 语音合成单元的时长等特性决定于其所处的上下文环境。对语音时长的预测, 本质上是从上下文环境参数的取值空间到时长取值空间的映射。本发明提出“逐步线性回归的时长预测模型”, 来逼近上述从上下文环境参数的取值空间到时长取值空间的映射。

[0144] 语音参数的生成是以语音合成单元的子状态为单位进行的，其先决条件之一是语音合成单元各子状态的时长，而逐步线性回归的时长预测模型生成的时长值只是具体到语音合成单元这一级别，所以需要将逐步线性回归的初始时长预测结果进行分配，获得语音合成单元各子状态的逐步线性回归的分配时长预测结果。

[0145] 在所述逐步线性回归的时长预测模型中，采用了“逐步参数优选”的策略，不仅可以为应对语料库对语音合成单元上下文环境的取值空间覆盖不足、不均衡等常见问题提供了一种有效的手段；还能够兼顾精简的同时，逐步真实地逼近从上下文环境参数的取值空间到时长取值空间的映射关系。

[0146] 所述“逐步参数优选”的策略是指，从众多的上下文环境参数中优选出对时长预测的贡献度较大者，从而有效提高预测模型的可外推性 (extrapolation) 和计算效率。逐步迭代的过程，从一个初始常参数开始，逐步引入待选参数中对预测误差的下降贡献最大者；重负这个过程，直至预测误差无下降或下降不明显为止。

[0147] 在模型训练过程中，估计模型参数和评价预测误差使用两组不同的样本集，辅之以参数集合的精简，可以有效地减小模型对训练数据的过度拟合，从而提高预测模型的可外推性。

[0148] 所述逐步线性回归的时长预测模型能够直接确定语音合成单元的持续时间，但是没有细分到其子状态层级，所以需要将新生成的时长返回到原模型中按照各子状态的比例等比例缩放，获得各子状态的持续时间，从而在下一步中确定各子状态的基频和谱参数的持续时间。在新方法中，既有的决策树 - 高斯混合模型只负责确定各子状态之间的比例分配，真正的时间长度只由新（回归）模型确定。

[0149] 在本发明的一种优选实施例中，所述采用决策树 - 混合模型对所述初始时长预测结果进行分配的步骤，可以进一步包括：

[0150] 针对上下文环境参数，采用决策树 - 高斯混合模型，对语音合成单元及各语音合成单元子状态的时长进行预测，获得语音合成单元时长和语音合成单元各子状态时长的缩放比例；

[0151] 根据语音合成单元时长和语音合成单元各子状态时长的缩放比例，将所述初始时长预测结果进行等比例缩放，获得语音合成单元各子状态的时长预测结果。

[0152] 参照图 4，示出了本发明一种语音合成的流程示意图，具体可以包括：

[0153] 步骤 401、输入需要进行语音合成的输入文本；

[0154] 步骤 402、对上述输入文本进行文本分析，提取出上下文环境参数；

[0155] 步骤 403、针对上述提取出的上下文环境参数，采用逐步线性回归的时长预测模型，对语音合成单元的时长进行初始预测，获得逐步线性回归的初始时长预测结果；

[0156] 步骤 404、采用决策树 - 高斯混合模型对所述初始时长预测结果进行分配，得到分配后的时长预测结果；

[0157] 步骤 405、依据分配后的时长预测结果，获得连续语音的参数的持续时间；

[0158] 其中，所述连续语音的参数，具体可以包括：语音合成单元各子状态的基频参数和谱参数；

[0159] 步骤 406、将所述连续语音的参数送入合成器，合成出语音。

[0160] 为使本领域技术人员更好地理解本发明，下面以一个具体的实例来说明上述通过

逐步线性回归模型预测出的时长值进行语音合成的步骤,具体可以包括:

- [0161] 子步骤 G1、输入需要进行语音合成的一句输入文本;
- [0162] 子步骤 G2、针对上述需要进行语音合成的一句输入文本进行文本分析,得到每个字的声母、韵母、声调,在所属的词、短语、句子中的位置,所属词、短语、句子的长度等信息,以及相邻字的信息;
- [0163] 子步骤 G3、针对上述需要进行语音合成的一句输入文本,采用逐步线性回归的时长预测模型,对语音合成单元的时长进行初始预测,获得逐步线性回归的初始时长预测结果;
- [0164] 子步骤 G4、将上述逐步线性回归的初始时长预测结果,进行分配,获得分配后的时长预测结果;
- [0165] 子步骤 G5、依据分配后的时长预测结果,计算出上述整句话的基频参数和频谱参数;
- [0166] 子步骤 G6、将上述整句话的基频参数和频谱参数送入合成器,合成出需要进行语音合成的一句输入文本的语音。
- [0167] 在本发明的另一优选实施例中,所述将上述逐步线性回归的初始时长预测结果,进行分配,获得逐步线性回归的分配时长预测结果的步骤,具体可以包括:
 - [0168] 子步骤 H1、对每个声 / 韵母,按照这些信息到其各子状态的决策树中查找,定位到具体某子类,从这些子类对应的决策树 - 高斯混合模型中计算获得时长值;
 - [0169] 子步骤 H2、根据子步骤 H1 中的时长值,确定每个声 / 韵母各子状态的重复次数;
 - [0170] 子步骤 H3、根据上述所确定每个声 / 韵母各子状态的重复次数,将逐步线性回归的初始时长预测结果进行分配,获得每个声 / 韵母各子状态的逐步线性回归的分配时长预测结果;
- [0171] 子步骤 H4、依据上述每个声 / 韵母各子状态的逐步线性回归的分配时长预测结果,获得每个声 / 韵母各子状态的基频参数和频谱参数。
- [0172] 在本发明的另一优选实施例中,所述依据逐步线性回归的分配时长预测结果,计算出上述整句话的基频参数和频谱参数的步骤,具体可以包括:
 - [0173] 子步骤 I1、使用与子步骤 H1 中类似的方法,定位到各声 / 韵母各个子状态的基频参数和频谱参数的子类,
 - [0174] 子步骤 I2、将所述各声 / 韵母各个子状态的基频参数和频谱参数的子类相应的子状态链以及定位的逐步线性回归的时长预测模型串接在一起;
 - [0175] 子步骤 I3、根据这个串接在一起的整体模型,计算出上述需要进行语音合成的一句输入文本的基频参数和频谱参数。
- [0176] 对于语音时长的预测方法实施例而言,由于其与训练方法实施例基本相似,所以描述的比较简单,相关之处参见训练方法实施例的部分说明即可。
- [0177] 参照图 5,示出了本发明一种逐步线性回归的时长预测模型的训练装置实施例的结构图,具体可以包括:
 - [0178] 建立模块 501,用于建立初始的线性回归的时长预测模型;及
 - [0179] 优化模块 502,用于在迭代所述线性回归的时长预测模型的过程中,通过评价每轮的时长预测模型选择上下文环境参数,最终得到最优时长预测模型。

[0180] 在本发明实施例中,优选的是,所述线性回归的时长预测模型的表达式如下:

$$\begin{aligned} [0181] \quad & \left\{ \begin{array}{l} Y = X\beta + \varepsilon \\ E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 \end{array} \right. \end{aligned}$$

[0182] 在本发明的一种优选实施例中,可以在迭代所述线性回归的时长预测模型的过程中,时长预测模型的估计和评价采用不同的样本。

[0183] 在模型训练过程中,估计模型参数和评价预测误差使用两组不同的样本集,辅之以参数集合的精简,可以有效地减小模型对训练数据的过度拟合,从而提高预测模型的(相对与训练集的)可外推性。

[0184] 对于训练系统实施例而言,由于其与训练方法实施例基本相似,所以描述的比较简单,相关之处参见训练方法实施例的部分说明即可。

[0185] 参照图6,示出了本发明一种语音时长的预测装置实施例的结构图,具体可以包括:

[0186] 初始时长预测模块601,用于针对上下文环境参数,采用逐步线性回归的时长预测模型,对语音合成单元的时长进行初始预测,获得初始时长预测结果;

[0187] 分配模块602,用于采用决策树-高斯混合模型对所述初始时长预测结果进行分配,得到分配后的时长预测结果。

[0188] 在本发明实施例中,优选的是,所述分配模块702可以进一步包括:

[0189] 子状态预测单元,用于针对上下文环境参数,采用决策树-高斯混合模型,对语音合成单元及各语音合成单元子状态的时长进行预测,获得语音合成单元时长和语音合成单元各子状态时长的缩放比例;

[0190] 缩放单元,用于根据语音合成单元时长和语音合成单元各子状态时长的缩放比例,将所述初始时长预测结果进行等比例缩放,获得语音合成单元各子状态的时长预测结果。

[0191] 对于语音时长的预测系统实施例而言,由于其与语音时长的预测方法实施例基本相似,所以描述的比较简单,相关之处参见语音时长的预测方法实施例的部分说明即可。

[0192] 本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0193] 以上对本发明所提供的一种逐步线性回归的时长预测模型的训练方法及装置、一种语音合成单元时长的预测方法及装置,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;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。

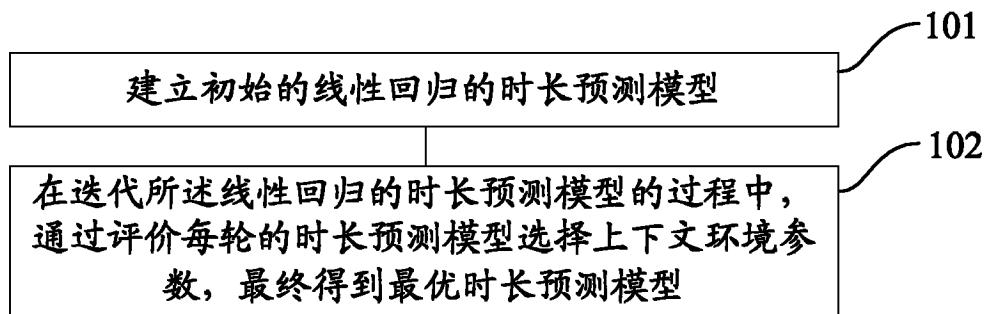


图 1

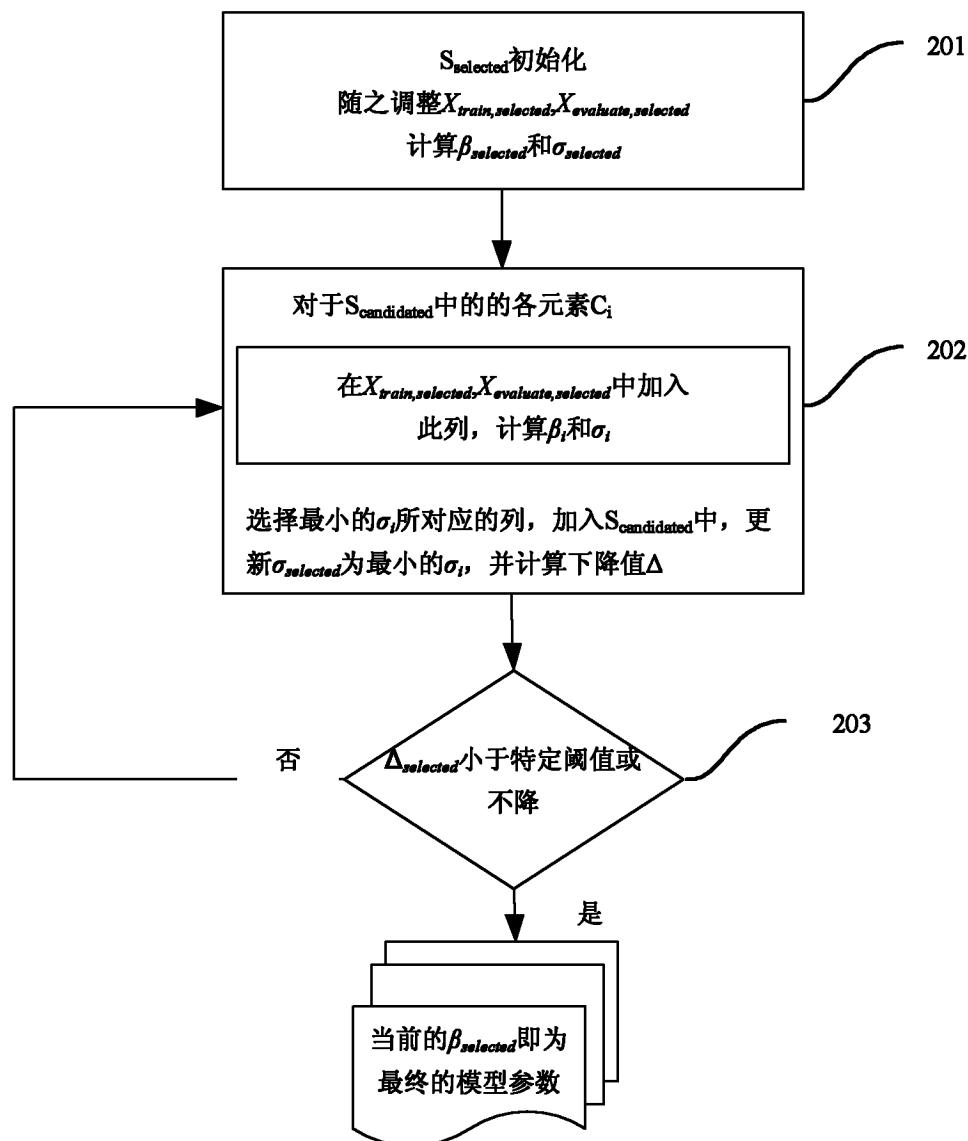


图 2

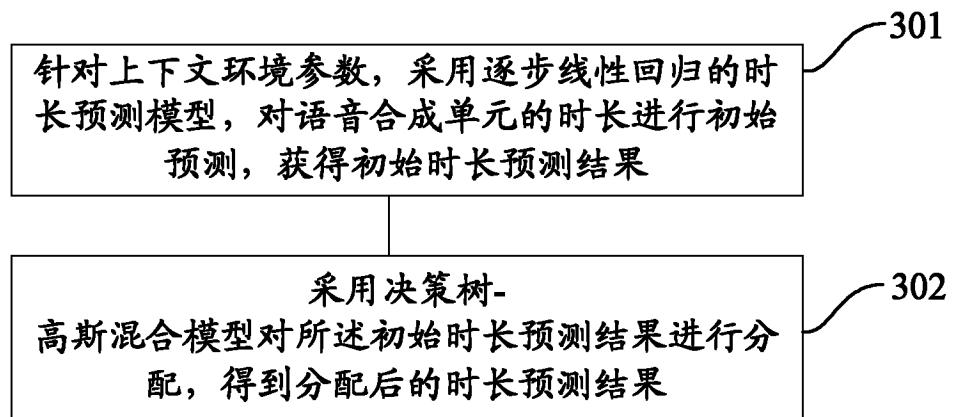


图 3

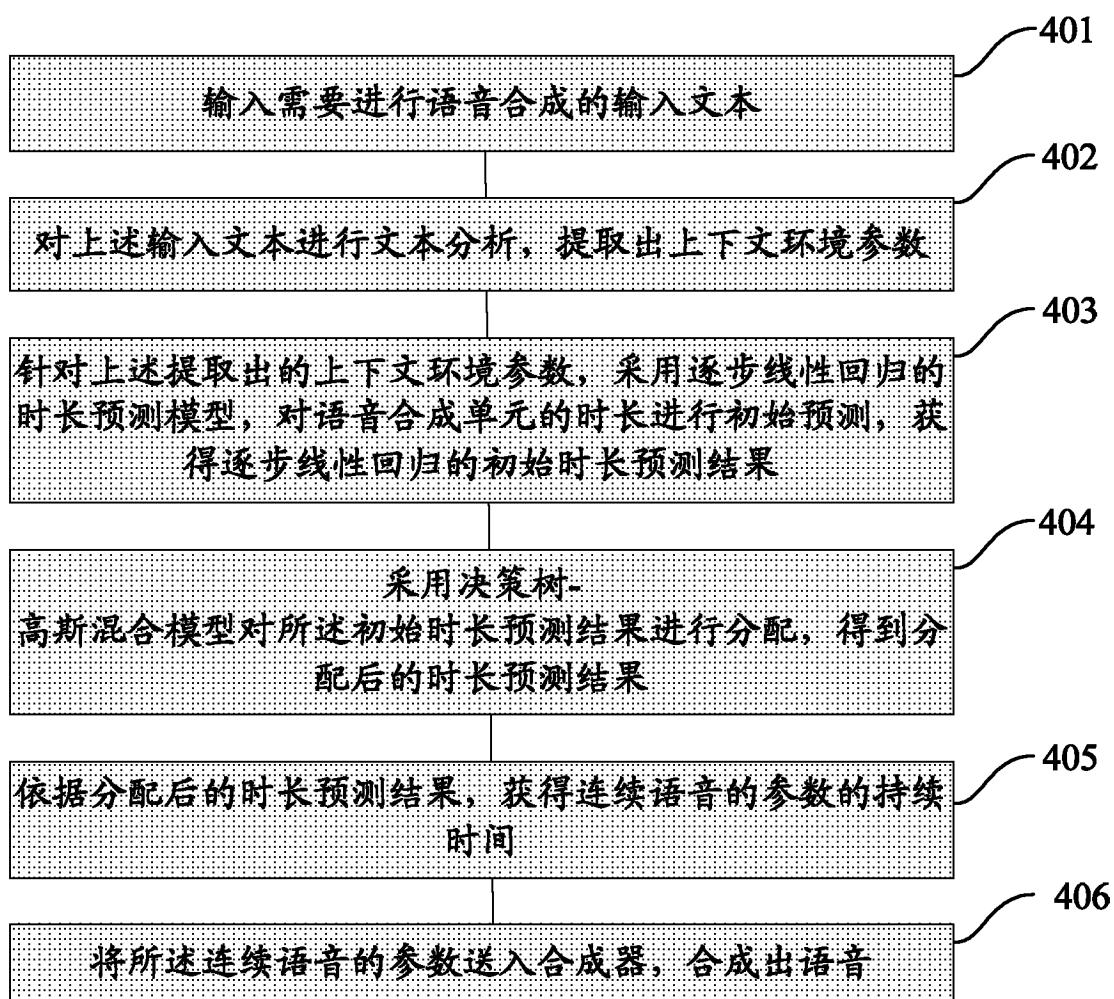


图 4

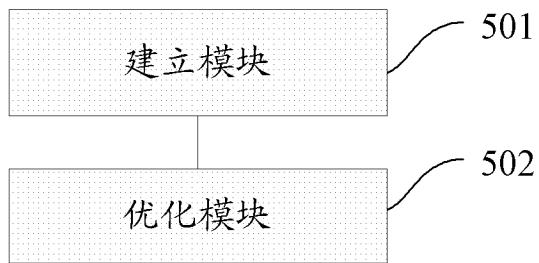


图 5

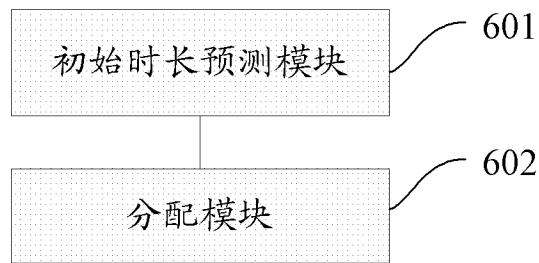


图 6