

doi:10.16055/j.issn.1672-058X.2022.0005.008

一种基于 HMM 算法改进的语音识别系统

黄清, 方木云

(安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243000)

摘要:针对传统的 HMM 模型中状态持续时间不长的不足,且在计算量大的情况下,语音识别精度不高,训练时间长,训练误差较高,提出了一种基于语音状态持续时间长的 HMM 模型。首先,令状态转移矩阵的对角线元素全为 0,去掉自转移弧,再增添以参数化的函数描述持续时间的高斯分布,再通过帧与帧相互之间的关联程度,将每帧都计算进去;其次,通过重估公式反复计算每条弧被指定的转变概率和可见符号序列输出最原始的数值概率,直至收敛,停止运算。最后,在 HMM 模型改进前后实验中得到更小的训练误差,下降速度更快,计算量较之前减少多,更容易达到收敛,其概率输出与它前面一个概率输出的差值与该概率输出值的比值大于 HMM 模型设定的初始值。与传统 HMM 模型实验比较,基于持续时间状态的 HMM 模型可以在一定程度上降低训练次数和训练时间,提高识别语音的精确度,基本完成了语音识别系统的功能。

关键词:语音识别;传统 HMM 模型;状态持续时间的 HMM

中图分类号:TP183

文献标志码:A

文章编号:1672-058X(2022)05-0056-06

1 相关工作

隐马尔可夫模型^[1](Hidden Markov Model, HMM)是一种统计信号模型,语音信号被分割成多个帧,将分割后的部门串起来,必须将这些部分串起来,这就需要用 Markov 链,而 HMM 模型就是通过 Markov 链演变形成的。HMM 是体现语音信号特征的动态变化非常合适的一种模型,在识别语音的过程中,其识别精准度是非常高,现在 HMM 已成为公认的描述语音信号最为有效的统计模型,在如今的许多产品中有着较为广泛的应用。

国外 2016 年 IBM 团队^[1]创造了 7.0% 的次错率新纪录,其解码部分采用的就是 HMM 模型,带有 maxout 激活的循环神经网络。微软 2019 年 12 月,利用了卷积以及 LSTM(Long Short Term Memory)神经网络,开创了一个全新的声学训练,将字词错误率降低到了近 5.9%。国内 2018 年,科大讯飞提出了深度全序列卷积神经网络,并采用了大量的卷积直接对整段

的语音信号进行建模处理,是为了获取语音信号中的信息。2019 年阿里提出将 LFR(Lancichinetti Fortunato Radicchi benchmark)和 DFSMN(Deep FSMN)算法进行融合^[2],语音识别错误率降低了 25%,解码速度提升了近 3 倍多。它们虽然都使语音识别的精度有所提高,但是语音识别率还不够精确。基于持续时间状态的 HMM 模型,确保原本的精度上,更加提高识别精确度。

如果一个语音信号包括时间短暂且具有稳定特性的,就可以借用 Markov 链来叙述它,并且可以及时跟踪他们之间的状态变化的情况。而仅仅通过它得出其输出的最终序列状态,却看不见它们之间的状态时间的改变情况和状态的分布信息,而 HMM 就是个双重随机过程,如图 1 所示。

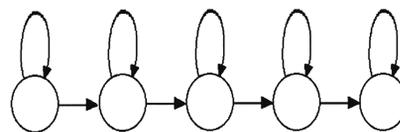


图 1 HMM 模型状态图

Fig. 1 HMM model state diagram

收稿日期:2021-04-22;修回日期:2021-06-07.

基金项目:国家自然科学基金项目资助(61971004).

作者简介:黄清(1995—),男,安徽枞阳人,硕士,从事高级软件工程研究.

通常用三组模型参数 $M = \{A, B, \pi\}$ 来定义用于语音识别的 HMM。假设某 HMM 有 N 个状态 $\{S_i\}_{i=1}^N$, 各参数的定义如下:

A 为状态转移概率矩阵, a_{ij} 是从状态 S_i 到状态 S_j 转移时的转移机率, 要符合 $0 \leq a_{ij} \leq 1$, $\sum_{j=1}^N a_{ij} = 1$ 这样的条件;

$$A = \begin{Bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{Bmatrix}$$

π 为系统初始状态概率的集合, $\{\pi_i\}_{i=1}^N$ 表示的是状态最原始情况, 是 s_i 的概率, 即

$$\pi_i = P[S_1 = s_i] \quad (1 \leq i \leq N) \quad \sum_{j=1}^N \pi_j$$

B 是一个含有可见输出值几率的集合。 $B = \{b_{ij}(k)\}$, 其中 $b_{ij}(k)$ 是从状态位 S_i 到状态为 S_j 转换过程中可见值 x_i 为 k 的输出几率。为了便于形象化展示这个模型, HMM 模型由马尔可夫链和随机过程组成的, 如图 2 所示。

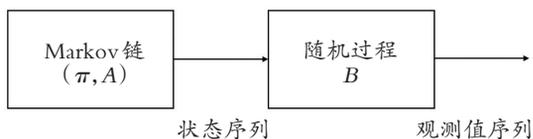


图 2 HMM 组成示意图

Fig. 2 Schematic diagram of HMM composition

2 隐马尔可夫模型的使用

在实际中要使用 HMM 模型来解决实际问题时, 就需要解决 3 个问题:

(1) 对于一个 HMM 模型, 如果已经有了观察值序列 $O = o_1, o_2, \dots, o_T$ 和模型 $M = \{A, B, \pi\}$, 怎么样来计算由模型 M 产生出 O 的概率 $P(O|M)$, 在识别语音的过程中, 需要计算得到 $P(O|M)$ 这个概率, 来寻找到与可观测值序列最高契合度的 HMM 模型。

(2) 在同样的观察值序列 $O = o_1, o_2, \dots, o_T$ 和模型 $M = \{A, B, \pi\}$, 如何找到一种状态转移的顺序, 有效得到与之对应的最佳状态序列。

(3) 怎样调整模型 $M = \{A, B, \pi\}$, 使得 $P(O|M)$ 最大。

2.1 计算输出概率 P

Forward-Backward 算法^[3]是按输出观察值序列的时间来通过一个特定观察序列 $O = o_1, o_2, \dots, o_T$ 以及一个模型 $M = \{A, B, \pi\}$ 时进行计算, 依靠模型 M 算出概率 $P(O|M)$ 。为了有效求出 $P(O|M)$, 鲍姆等人提出前后向算法。设状态的起始值 S_1 , 状态的结束值是 S_N , 以下为前后向算法:

(1) 前向算法。该算法凭借可见序列的检测值依次递进推出输出的计算概率。 $O = o_1, o_2, \dots, o_T$ 为可见信号的输出序列, $P(O|M)$ 为模型 M 可见信号的输出序列 O 的概率, a_{ij} 从 S_i 到 S_j 变化状态的概率, $b_{ij}(o_t)$ 从 S_i 到 S_j 转变时输出状态 o_t 的概率, $\alpha_i(j)$ 为部分信号输出序列 o_1, o_2, \dots, o_i 并且转换至 S_j 的状态概率, 称之为前向率, 前向率 $\alpha_i(j)$ 是依靠下面的公式推导计算得到:

初始化过程

$$\alpha_0(1) = 1, \alpha_0(j) = 0 (j \neq 1)$$

递推过程

$$\alpha_t(j) = \sum_i \alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \quad (t = 1, 2, \dots, T, i, j = 1, 2, \dots, N)$$

最终输出结果

$$P(O|M) = \alpha_T(N)$$

在运用递推前向算法过程中计算模型 $M = \{A, B, \pi\}$ 在输出观察符号序列为 o_1, o_2, \dots, o_T 时的输出概率 $P(O|M)$ 的具体步骤:

① 将每一个状态分配一个可变量数组 $\alpha_i(j)$, 在初始化进程时设置状态 S_1 的最初变量数组 $\alpha_0(1)$ 为 1, 状态的其余变量数组 $\alpha_0(j)$ 为 0。

② 根据 t 时间的可见信号 o_t 的输出值情况, 计算出 $\alpha_i(j)$:

$$\alpha_t(j) = \sum_i \alpha_{t-1}(i) a_{ij} b_{ij}(o_t) = \alpha_{t-1}(1) a_{1j} b_{1j}(o_t) + \alpha_{t-1}(2) a_{2j} b_{2j}(o_t) + \dots + \alpha_{t-1}(N) a_{Nj} b_{Nj}(o_t) \quad (j = 1, 2, \dots, N)$$

当状态 S_i 到状态 S_j 没有转移时, $a_{ij} = 0$ 。

③ 当符合 $t \neq T$ 这个条件时跳转到第二步, 相反就进行第四步。

④ 把最新运算得到的变量数组中 $\alpha_T(N)$ 的所以数值都提取出来, 则 $P(O|M) = \alpha_T(N)$

(2) 后向算法。设定 $\beta_t(i)$ 是后向率, 从设 S_i 为最初状态到设 S_N 为终止状态过程中少许信号序列输出的概率为 $o_{i+1}, o_{i+2}, \dots, o_T$, 则 $\beta_t(i)$ 可借助下列公式递推计算获得:

初始化过程

$$\beta_T(N) = 1, \beta_T(j) = 0, (j \neq N)$$

递推过程

$$\beta_t(i) = \sum_j \beta_{t+1}(j) a_{ij} b_{ij}(o_{t+1}) \quad (t = T, T-1, \dots; i, j = 1, 2, \dots, N)$$

最终输出结果

$$P(O|M) = \sum_{i=1}^N \beta_1(i) \pi_i = \beta_0(1)$$

根据定义的前后向概率, 则有如下等式成立:

$$P(O|M) = \sum_{i=1}^N \sum_{j=1}^N \alpha_i(i) a_{ij} b_{ij}(o_{i+1}) \beta_{i+1}(j),$$

$$1 \leq t \leq T - 1$$

2.2 求状态转移序列

维特比算法^[3]处理的是给定一个可见值序列 $O = o_1, o_2, \dots, o_T$ 和一个特定模型 $M = \{A, B, \pi\}$ 时, 准确得到一个程度最合适的状态序列 $S = s_1, s_2, \dots, s_T$ 的问题。状态序列程度最合适的意思是指先让 $P(S, O|M)$ 达到最大时得到的状态最准确序列。

初始化过程

$$\alpha'_0(1) = 1, \alpha'_0(j) = 0 (j \neq 1)$$

递推过程

$$\alpha'_t(j) = \max_i \alpha'_{t-1}(i) a_{ij} b_{ij}(o_t) \\ (t = 1, 2, \dots, T; i, j = 1, 2, \dots, N)$$

最终输出结果

$$P_{\max}(S, O|M) = \alpha'_T(N)$$

在这个递推过程当中, 形成的状态序列中 $\alpha'_t(j)$ 都是状态 i 最大的就是所求的状态序列最理想的情况。求取状态序列最理想化运用维特比算法推导步骤如下:

① 将每一个状态分配一个可变量的数组 $\alpha'_t(j)$, 在初始化阶段时设置状态 S_1 最初变量数组 $\alpha'_0(1)$ 为 1, 状态的其余变量数组 $\alpha'_0(j)$ 为 0。

② 根据 t 时间的可见信号 o_t 的输出值情况, 计算出 $\alpha'_t(j)$

$$\alpha'_t(j) = \max_i \alpha'_{t-1}(i) a_{ij} b_{ij}(o_t) = \max_i \alpha'_{t-1}(1) a_{1j} b_{1j}(o_t) + \\ \max_i \alpha'_{t-1}(2) a_{2j} b_{2j}(o_t) + \dots + \\ \max_i \alpha'_{t-1}(N) a_{Nj} b_{Nj}(o_t) (j = 1, 2, \dots, N)$$

当状态 S_i 到状态 S_j 没有转移时, $a_{ij} = 0$ 。设计一个符号数组变量, 称为最佳状态序列寄存器, 利用这个最佳状态序列寄存器把每一次使 $\alpha'_t(j)$ 最大的状态 i 保存下来。

③ 当符合 $t \neq T$ 这个条件时跳转到第二步, 相反就进行第四步操作。

④ 把最新的变量数组中 $\alpha'_T(N)$ 的所有数值都提取出来, 则 $P_{\max}(S, O|M) = \alpha'_T(N)$ 为序列状态最理想化的寄存器中存留的值, 即为所求的状态序列程度最合适。

2.3 模型参数训练

Baum-Welch 算法^[4]的实质是为了完成 HMM 训练集中参数的估计问题的解决。在特定一个可见值序列 $O = o_1, o_2, \dots, o_T$ 以及一个模型 $M = \{A, B, \pi\}$ 时, 使得 $P(O|M)$ 达到最大值。在这里, 需要将 M 求出来才能让 $P(O|M)$ 值成为最大, 但是由于受到训练序列的给定限制, 因此无法把 M 估计值达到最理想的地步, 而含有递归的思想的 Baum-Welch 算法就把问题简化了, 先将 $P(O|M)$ 中的一部分扩大, 最后算出的模型参数 $M = \{A, B, \pi\}$ 就是经过优化之后的, 然后

再借助 Baum-Welch 算法中的公式得到的重新估算后的 \hat{M} 。将该过程循环进行, 一步步修改模型优化参数, 当 $P(O|\hat{M})$ 不再很显著增长达到收敛的效果, 这个时候的 \hat{M} 就是所希望得到的理想模型。

在时间为 t 时 S_i 转换至 S_j 时的稳态概率是 $\gamma_t(i, j)$, 则 $\gamma_t(i, j)$ 的公式定义为

$$\gamma_t(i, j) = \frac{\alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \beta_t(j)}{\alpha_t(N)} = \frac{\alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \beta_t(j)}{\sum_i \alpha_t(i) \beta_t(i)}$$

同时, 对于信号字符序列 $O = o_1, o_2, \dots, o_T$, Markov 链在时间为 t 时存在的 S_i 状态的输出概率公式为

$$\sum_{j=1}^N \gamma_t(i, j) = \frac{\alpha_t(i) \beta_t(i)}{\sum_i \alpha_t(i) \beta_t(i)}$$

对信号字符序列 $O = o_1, o_2, \dots, o_T$, 其转变状态次数为 $\sum \gamma_t(i, j)$ 的期待值是从 S_i 转变至 S_j 的数值; 而从 S_i 再输出去状态转变次数的期待值为 $\sum_j \sum_t \gamma_t(i, j)$ 。这里可借用 Baum-Welch 算法中导出公式重新估计值:

$$\hat{a}_{ij} = \frac{\sum_t \gamma_t(i, j)}{\sum_j \sum_t \gamma_t(i, j)} = \frac{\sum_i \alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \beta_t(i)}{\sum_i \alpha_t(i) \beta_t(i)} \quad (1)$$

$$\hat{b}_{ij}(k) = \frac{\sum_{t: o_t=k} \gamma_t(i, j)}{\sum_t \gamma_t(i, j)} = \frac{\sum_{t: o_t=k} \alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \beta_t(j)}{\sum_i \alpha_{t-1}(i) a_{ij} b_{ij}(o_t) \beta_t(j)} \quad (2)$$

故依靠可见值序列 $O = o_1, o_2, \dots, o_T$ 和模型 $M = \{A, B, \pi\}$ 被选择出来的原始状态, 由导出公式重新估计式(1)和式(2), 求出的参数 \hat{a}_{ij} 和 $\hat{b}_{ij}(k)$ 是一组全新的数值^[5], 一个崭新的重新估计模型 $\hat{M} = \{\hat{A}, \hat{B}, \hat{\pi}\}$ 就这样横空出世了。

3 隐马尔可夫训练过程的改进和仿真比较

3.1 改进 HMM 的训练过程

传统的 HMM 描述状态持续时间具有较大的局限性, 针对此问题, 最直接的方法就是非参数方法, 即在传统的 HMM 中, 令状态转移矩阵 A 的对角线元素全为 0 ($a_{ii} = 0$) 则各状态没有自转移弧, 再增添状态分布 $p_i(d)$, 设 D 是所有状态当中停留时间最长的, 考虑状态持续时间的因素, HMM 用 Viterbi 算法进行语音识别时, 考虑状态持续时间的因素的 HMM 的输出概率为

$$\log_{10} a_n(j) = \max \left\{ \max \left[\log_{10} a_{n-\tau}(i) + \log_{10} a_{ij} + \omega \log_{10} d_j(\tau) + \sum_{k=1}^{\tau} \log_{10} b_{ij}(Y_{n-k+1}) \right] \right\}$$

从而在传统的 HMM 中,将自转移弧去掉,用 τ 表示自转移次数即停留帧数,并用 $\sum_{k=1}^{\tau} \log_{10} b_{ij}(Y_{n-k+1})$ 将每帧均计算进去,通过帧与帧相互之间的关联程度来进行分类分段的,将信号语音的特征体现得更加淋漓尽致,以此为基础,改进的 HMM 参数的重估公式为

$$\begin{aligned} \hat{\pi} &= \frac{\pi_i \hat{\beta}_n(i)}{P(Y| \lambda)} \\ \hat{a}_{ij} &= \frac{\sum_{i=1}^N a_n(i) a_{ij} \hat{\beta}_n(j)}{\sum_{j=1}^N \sum_{i=1}^N a_n(i) a_{ij} \hat{\beta}_n(j)} \\ \hat{b}_{jk} &= \frac{\sum_{n=1}^N \left[\sum_{\tau < n} \hat{a}_{\tau}(j) \hat{\beta}_{\tau}(j) - \sum_{\tau < n} \alpha_{\tau}(j) \beta_{\tau}(j) \right]}{\sum_{k=1}^M \sum_{n=1}^N \left[\sum_{\tau < n} \hat{a}_{\tau}(i) \hat{\beta}_{\tau}(j) - \sum_{\tau < n} \alpha_{\tau}(j) \beta_{\tau}(j) \right]} \\ \hat{P}_i(d) &= \frac{\sum_{n=1}^N \hat{a}_{\tau}(i) \rho_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} b_i(y_s)}{\sum_{d=1}^D \sum_{n=1}^N \hat{a}_{\tau}(i) \rho_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} b_i(y_s)} \end{aligned}$$

其中, π_i 的估计为状态 i 作为序列第一个状态的概率, \hat{a}_{ij} 的估计和传统的 HMM 类似, \hat{b}_{jk} 为状态 S_j 下, 观察矢量为 $y_n = V_k$ 发生的次数与发生转移的全部

次数之比。 $\hat{P}_i(d)$ 的重估是在任何持续长度下, 模型处于状态 S_j 的比率。增加了以状态持续时间为参数的 HMM 模型更为精确, 比传统 HMM 有更好的语音识别性能, 但计算量和存储增加, 特别是估计可靠的参数需要很大的训练数据量。为了降低训练中的计算量, 采用以参数化的函数描述持续时间的分布函数, 即高斯分布, $\rho_i(d) = N(d, \mu_i \sigma_i^2)$ 。

3.2 仿真实验比较

实验采用的是 0 到 9 的阿拉伯数字^[6], 因为与其他的语种相比, 数字更加可以验证实验的结论。因为它们之间的发音有太多地方是相同的, 并且识别数字的难易程度更高。为了比较传统的 HMM 算法与改进方法的训练效果, 取了不同时间段的语音分为两组, 这里设 HMM 的状态为 4, HMM 算法仿真中, 先给定的一个 HMM 模型初始值如表 1 所示。

表 1 HMM 模型初始值

分 组	1	2	3	4
1	0.450 0	0.450 0	0	0
2	0	0.450 0	0.450 0	0
3	0	0	0.450 0	0.450 0
4	0	0	0	1

对传统 HMM 再进行对照训练, 如表 2 所示的就是信号语音的训练过程, 多次的反复训练之后, 当其概率输出与它前面一个概率输出的差值与该概率输出值的比值大于 HMM 模型初始值 4.5×10^{-6} 。训练之后的 HMM 模型的参数如表 3 所示。

表 2 传统 HMM 模型训练过程

Table 2 Traditional HMM model training process

训练次数	1	2	...	23	24
输出概率 (log) = -6.230 571e+003	(log) = -6.219 725e+003	...	(log) = -6.171 801e+003	(log) = -6.171 811e+003	收敛

表 3 传统 HMM 模型训练结果

Table 3 Training results of traditional HMM model

分 组	1	2	3	4
1	0.857 1	0.542 3	0	0
2	0	0.825 9	0.460 7	0
3	0	0	0.835 2	0.550 4
4	0	0	0	1

两个办法各自的训练次数都为 24, 得到其对比仿真图如图 3 所示。图 3 仿真结果表明, 改进后的 HMM 训练 4 次后, 就可以达到较小的训练误差, 并且训练收敛速度较快。而传统的 HMM 训练 8 次以上才能得到较小的训练误差, 故 HMM 模型通过优化后的训练较之前下降得更快, 更方便实现收敛, 在一定意义上突出了改进办法的优越性。图 4 是训练后的识别结果。

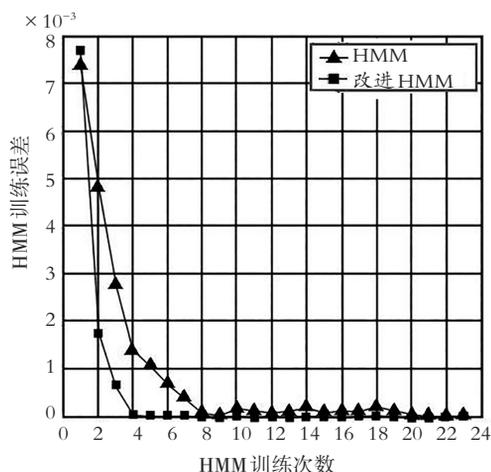


图 3 改进前后对 HMM 模型训练的性能影响

Fig. 3 The performance impact of HMM model training before and after the improvement

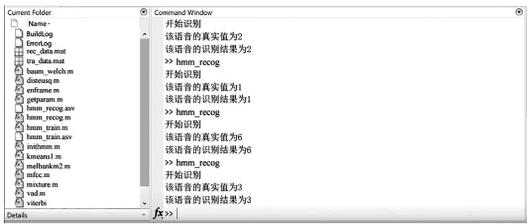


图 4 HMM 算法识别结果

Fig. 4 Recognition results of HMM algorithm



图 7 机器人控制功能图

Fig. 7 Robot control function diagram

4 系统功能的实现

完成了各个部分的设计和编程后,就进行整个系统的效果检验^[7]。本文所用的语音信号的都是预处理之后得出的语音信息,但首先需初始化客户端,与服务端建立连接。

(1) 语音识别功能。在语音识别模块中,先选取之前已经录制好的特定工作人员的语音,然后点击识别按钮(图 5),就会将录制好的语音信号很准确地翻译出来,并且可以判断这段录制好的语音文件是一个有效的波形文件,增加了语音识别系统的可信度^[8]。



图 5 语音识别功能图

Fig. 5 Diagram of speech recognition function

(2) 机器翻译功能。在将语音信号正确识别出来之后,该模块将语音指令转化为机器人指令。下图为一个语音指令成功转化^[9]为机器人指令“停止工作”的图(图 6)。

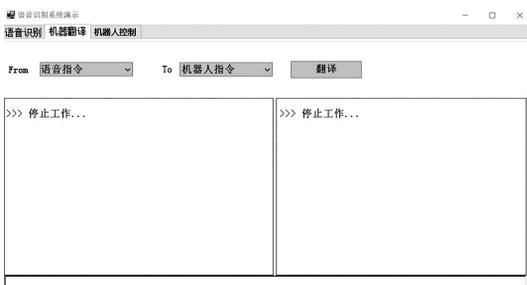


图 6 机器翻译功能图

Fig. 6 Machine translation function diagram

(3) 机器人控制功能。该模块中有开始训练,手动控制,语音控制,结束功能,图 7 是点击开始训练,语音控制的效果图^[10]。

5 结束语

主要介绍了 HMM 模型的概念,包括隐马尔可夫的定义和其 3 个问题的解决算法,其中重点介绍了 HMM 算法,最初的方式训练过程,基于信号语音的特点,对优化 HMM 模型的训练,再借助实验仿真的结果,证实了该方法具有优越性,训练改进后的 HMM 模型次数与时间均有所降低。并最终在 HMM 改进算法基础上实现了机器人语音识别系统。

参考文献 (References):

[1] 刘畅,魏忠诚,张春华,等. 基于隐马尔可夫模型的步态识别算法[J]. 计算机工程与设计, 2019, 40(12): 3487—3493.
 LIU Chang, WEI Zhong-cheng, ZHANG Chun-hua, et al. Gait recognition algorithm based on hidden Markov model[J]. Computer Engineering and Design, 2019, 40(12): 3487—3493.

[2] 赵明明,王洪春. 一种用于语音识别的高效分帧函数的研究[J]. 重庆工商大学学报(自然科学版), 2012, 29(5): 68—72.
 ZHAO Ming-ming, WANG Hong-chun. Research on an efficient framing function for speech recognition [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2012, 29(5): 68—72.

[3] 李锦,何培宇. 一种改进的基于小波去噪 HMM 非特定人语音识别算法[J]. 四川大学学报(自然科学版), 2007(1): 69—72.
 LI Jin, HE Pei-yu. An improved wavelet denoising HMM non-specific speech recognition algorithm [J]. Journal of Sichuan University (Natural Science Edition), 2007(1): 69—72.

[4] 许丽红,余小清,万旺根. 基于 HMM 算法的语音识别系统设计及其混合编程实现[J]. 计算机工程, 2002(5): 139—140, 163.
 XU Li-hong, YU Xiao-qing, WAN Wang-gen. Design of speech recognition system based on HMM algorithm and its hybrid programming implementation [J]. Computer Engineering, 2002(5): 139—140, 163.

- [5] 何强,毛士艺,张有为.汉语语音识别的平滑声韵基元HMM算法[J].北京航空航天大学学报,2001(2):146—149.
HE Qiang, MAO Shi-yi, ZHANG You-wei. Smooth phonological primitive HMM algorithm for Chinese speech recognition[J]. Journal of Beijing University of Aeronautics and Astronautics, 2001(2):146—149.
- [6] 张起浩,将少国,赵鹏.HMM算法语音识别控制机械手的系统设计[J].单片机与嵌入式系统应用,2019,19(10):82—85.
ZHANG Qi-hao, JIANG Shao-guo, ZHAO Peng. System design of speech recognition control manipulator based on HMM algorithm[J]. Application of MCU and Embedded System, 2019,19(10):82—85.
- [7] 李强,陈浩,陈丁当.基于隐马尔可夫模型的语音激活检测算法[J].计算机应用,2016,36(11):3212—3216.
LI Qiang, CHEN Hao, CHEN Ding-dang. Speech activation detection algorithm based on hidden Markov model[J]. Computer Applications, 2016, 36(11):3212—3216.
- [8] 高珍珍,鲍长春.能量匹配的MFS-HMM语音增强方法[J].信号处理,2016,32(8):937—944.
GAO Zhen-zhen, BAO Chang-chun. MFS-HMM speech enhancement method based on energy matching[J]. Signal Processing, 2016,32(8):937—944.
- [9] 赵硕.基于HMM的非特定人语音识别系统研究[D].合肥:安徽大学,2019.
ZHAO Shuo. Research on unspecified person speech recognition system based on HMM[D]. Hefei: Anhui University, 2019.
- [10] 周楠,艾剑良.基于HMM和RNN的无人机语音控制方案与仿真研究[J].系统仿真学报,2020,32(3):464—471.
ZHOU Nan, AI Jian-liang. Research on UAV voice control scheme and simulation based on HMM and RNN[J]. Journal of System Simulation, 2020,32(3):464—471.

An Improved Speech Recognition System Based on HMM Algorithm

HUANG Qing, FANG Mu-yun

(School of Computer Science and Technology, Anhui University of Technology, Anhui Maanshan 243000, China)

Abstract: In view of the shortage of the traditional HMM model with a short state duration, and the low accuracy of speech recognition, long training time and high training error in the case of large computation, a HMM model based on a long state duration of speech was proposed. First, the diagonal elements of the state transition matrix are all 0, the self-transition arc is removed, and a Gaussian distribution describing the duration with a parameterized function is added. Then, each frame is calculated according to the degree of correlation between frames, and the specified transition probability of each arc and the most primitive numerical probability of the visible symbol sequence output are repeatedly calculated by the re-evaluation formula until convergence, and the operation is stopped. The ratio of the difference between its probability output and its previous probability output and the probability output value is greater than the initial value set by the HMM model. Compared with the traditional HMM model experiment, the HMM model based on the duration state can reduce the number of training times and shorten the training time to a certain extent, improve the accuracy of speech recognition, and basically complete the function of the speech recognition system.

Key words: speech recognition; traditional HMM model; HMM with state duration

责任编辑:陈芳

引用本文/Cite this paper:

黄清,方木云.一种基于HMM算法改进的语音识别系统[J].重庆工商大学学报(自然科学版),2022,39(5):56—61.

HUANG Qing, FANG Mu-yun. An improved speech recognition system based on HMM algorithm[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022,39(5):56—61.