

基于HMM方法的动态手势识别技术的改进

于美娟 马希荣

(天津师范大学计算机与信息工程学院 天津 300387)

摘要 为了提高基于HMM方法的动态手势识别的效率和准确性,针对HMM方法在训练手势中计算的高复杂性,提出了一种HMM算法和动态规划的算法相结合的方法,对HMM算法中的训练阶段进行了改进,增强了人机交互的准确性与实时性。

关键词 动态手势识别,隐马尔可夫模型,动态规划

Improvement of Dynamic Hand Gesture Recognition Technology Based on HMM Method

YU Mei-juan MA Xi-rong

(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China)

Abstract In order to enhance the efficiency and the accuracy of dynamic hand gesture recognition based on HMM method, for the computation's high complexity of HMM method in the training stage, a new HMM algorithm based on the dynamic programming was presented. It makes improvement to the HMM algorithm's training stage, and will enhance the accuracy and timeliness in human-robot interaction.

Keywords Dynamic hand gesture recognition, HMM, Dynamic programming

1 引言

人与计算机的交互活动越来越成为人们日常生活的一个重要组成部分,特别是最近几年,随着计算机技术的迅猛发展,研究符合人机交流习惯的新颖人机交互技术变得异常活跃,也取得了可喜的进步。

手势是一种自然而直观的人机交流模式。在人机交互中,手势识别是实现新一代人机交互所不可缺少的一项关键技术,所谓手势识别简单来理解就是指计算机根据手和胳膊的运动轨迹识别出它的含义,其包括静态手势识别(指姿态,单个手形)和动态手势(指动作,由一系列姿态组成)。相比之下,静态的手势识别比较简单,目前常用的手势识别技术有模板匹配法、神经网络法和统计分析法等,对于静态手势的识别这些技术的准确性还是不错的,但对于动态手势的识别,识别精度有待改进,因为动态手势是时变信号,与语音识别类似,同属动态连续信号的处理,所以近些年来有很多人把在语音识别中效果良好的HMM方法应用到了动态手势识别中。

2 HMM模型

隐马尔可夫(HM)过程是一个双重的随机过程,它包括状态转移的随机过程和观察值输出的随机过程。其中状态转移的随机过程是隐式的,通过观察序列的随机过程表现出来。人的手势可表示成双随机过程,尤其是内部状态不能直接观察到的Markov过程。对于这种过程中每一状态的转移,将产生其值依赖于各个状态的观察输出信号。因此,手势信号

可由参数随机过程来很好地刻画,并且该随机过程的参数可以用精确的、已定义好的方式来确定。用HMM来模型化手势,使得处理手势行为的高度随机性成为可能。

一个HMM的描述包括3个要素:

(1)一定数目($N+1$)的状态 $S = \{S_0, S_1, S_2, \dots, S_N\}$, 以及初始时刻 $t=0$ 时处于状态 S_i 的概率 $\pi_i = P(q(0)=S_i)$ 。

(2)状态的转移概率

$$P(q(t)=S_j, q(t-1)=S_i) = a_{ij}, i, j = 0 \dots N$$

描述了在 $t-1$ 时刻(前一时刻)处于状态 S_i , 而在 t 时刻(当前时刻)进入状态 S_j 的概率。一般记 $A = \{a_{ij}\}$ 。

(3)观察值的输出概率

$P(y(t)=k, q(t)=S_j) = b_{jk}$, 描述了在 t 时刻, 处于状态 S_j 而输出观察值 k 的概率。一般记 $B = \{b_{ij}\}$ 。

在比较完整的定义^[1]中将一个HMM的完整参数集合表示为 $\lambda = (M, N, \pi, A, B)$, 其中 N 表示模型中马尔可夫链的状态数目, M 表示每个状态对应的可能的观察值数目。在这里我们根据上文描述的HMM的三要素, 使用比较简单的表示法 $\lambda = (\pi, A, B)$ 。HMM在实际应用时需要解决3个基本问题: 概率的计算、最佳状态链的确定(Viterbi算法)、HMM参数优化问题, 本文中主要解决的是第三个问题。

3 HMM在动态手势识别中的应用

3.1 HMM在动态手势识别中的应用

HMM算法已经在手势识别方面有了比较广泛的应用, 例如Schlenzig等人^[5]将HMM用于视觉手势识别, 他们为每

到稿日期: 2010-03-12 返修日期: 2010-05-19 本文受国家自然科学基金项目(60970060/F020508), 天津市科技支撑重点项目(09ZCKFGX00500)资助。

于美娟(1985-), 女, 硕士生, 主要研究方向为模式识别和情感计算, E-mail: amy_1121@163.com; 马希荣(1962-), 女, 博士后, 教授, 主要研究方向为和谐人机交互、人工智能。

个手势定义了一个 HMM,该研究表明了 HMM 在连续手势识别中应用的能力。还有 Nam 与 Wohn^[6] 的时空手势识别系统提出了基于 HMM 用于识别时空手势运动模式的方法。美国 CMU 的 Christopher Lee 与 Yangsheng Xu^[7] 开发了一个基于 HMM 的手势识别系统等。

使用 HMM 方法训练手势的基本步骤^[1]如下:

(1)隐马尔可夫模型初始化:状态转移概率矩阵 A 的初始化,设定状态 i 只能返回到本身或转移到 $j=i+1$ 状态,即初始化一个 HMM 模型 $\lambda=(\pi, A, B)$ 。

(2)采用 Viterbi 算法计算出观察向量 O 在这个模型下的 $P(O|\lambda)$:初始模型确定以后,利用 Baum-Welch 算法对初始 HMM 模型进行重新计算,HMM 的各个参数在这一过程得到重新估计,得到新的 $\lambda=(\pi, A, B)$ 。利用 Viterbi 算法计算出观察序列 O 在这个模型下的 $P(O|\lambda)$ 。重复计算直至 P 收敛,即得到训练出的 HMM 和最接近观察值的隐马尔可夫模型。

我们可以看出 HMM 在动态手势识别中的应用,主要需要解决训练问题和评估问题。模型的训练问题是用给定训练手势集来确定模型参数 $\lambda=(A, B, \pi)$,使得 $P(O|\lambda)$ 达到局部最大。如上采用 Baum-Welch 算法,按照最大似然准则来优化模型参数。评估问题是观察序列与模型的匹配问题,即识别问题。假定识别系统的词表共包括 V 个词条。对从手势输入设备采集到的原始数据进行预处理,产生观察序列 $O=O_1, O_2, \dots, O_T$ 。孤立词的识别是针对每个动态手势词模型 $\lambda_i=(A_i, B_i, \pi_i)$ 计算观察序列的概率 $P(O|\lambda_i)$ ($0 \leq i \leq V$),取其中最大的作为识别结果,一般采用前向-后向(forward-backward)算法^[8]。

3.2 动态手势识别中 HMM 方法的不足

虽然 HMM 方法在动态手势识别方面已经有了比较广泛的应用,但由于不同的动态手势包含的基本手势词的数目是不同的,因此在基于 HMM 方法的动态手势识别中,对于同一个手势,建立不同的 HMM 状态模型,其识别实时性和准确性是不同的。从上面的论述中我们可以看出 HMM 识别算法在训练阶段需要提供大量的训练数据,通过反复计算才能得到模型参数,这属于一种穷举搜索法,计算量大,效率不高,尤其不太适合实时的手势识别。另外,在现实生活中,不同的人在做相同的手势或者同一人在不同的时间做相同的手势时,延续时间长度不同,单纯地使用 HMM 方法,在人机交互中并不能保证动态手势识别的准确性,所以为了增强动态手势识别的高效性和高准确度,我们引入动态规划的思想,提出一种基于动态规划算法的 HMM 方法,对基于 HMM 算法的动态手势识别进行改进,寻找一条最优匹配路径,在 HMM 的数据训练阶段做一定的变换,使 HMM 的评估函数尽快收敛,以建立适于不同动态手势的 HMM 状态模型。

3.3 对 HMM 方法的改进

动态规划理论是一种重要的算法设计思想,具有广泛的应用价值。使用动态规划来设计算法,对于不少问题往往具有高时效性,因而,对于能够使用动态规划来解决的问题,使用动态规划是比较明智的选择。动态规划算法的基本思想是采用多阶段决策策略,将一个 N 维问题转换为 N 个一维问题,对问题进行分级处理。

用动态规划算法求解实际问题时,首先要建立动态规划模型,一般需要进行如下几方面工作:

- (1)找出最优解的性质,并刻画其结构特征。
- (2)递归地定义最优值。

(3)以自底向上的方式计算出最优值,构造最优解。

用动态规划解决具体问题一般有顺序和逆序两种解法:

①以问题的初始状态为始端,结束状态为终端的解法为顺序解法;②以问题的初始状态为终端,结束状态为始端的解法为逆序解法。本文提出的改进方法采用的是逆序的方法。

有人在一些手语识别的研究中提出将动态时间规整算法 DTW 和 HMM 结合使用^[4],采用双层分类器的思想,使用 DTW 算法对收集到的动态手势初始数据和手势词表进行预处理,从宏观上对收集到的数据进行分类,然后使用二级 HMM 识别器,对手势进行最终详细的识别和理解。

本文采用将动态规划的思想融入到 HMM 中的方法,主要的目的是降低 HMM 算法在训练阶段的计算复杂度,提高效率。首先将收集到的原始动态手势数据按时间特性划分成两个阶段,简单地可以看成 HMM 状态模型中的两个状态 $O=\{O_1, O_2\}$,初始化 $\lambda=(A, B, \pi)$,对状态划分进行简单的调节,使得 $P(O|\lambda)$ 当前最大,因为当前只有两个状态, λ 的变换和 $P(O|\lambda)$ 的计算都比较简单;按照最优子结构的性质,原问题的最优解包含子问题的最优解,下面采用递归方法将 O_1, O_2 两个阶段逐步划分,每次以局部的 $P(O|\lambda)$ 最大为标准,分别计算两个阶段的最优解,对动态手势进行逐步划分,直到构造出最优的 HMM 状态模型,并与模板库中的数据序列匹配得到识别结果。

结束语 本文针对基于 HMM 的动态手势识别中训练阶段计算的复杂性,利用 HMM 状态分割中最优子结构的性质,将动态规划的思想与 HMM 相结合,采用自底向上的逆序方法,递归地建立应用于 HMM 识别的最优状态模型,它与单纯地使用 HMM 方法识别手势的方法相比,在训练阶段因为每次训练的状态数少,计算简单了很多,在状态的调整和似然函数的收敛上都能提高效率,相当于在整体上缩小了搜索范围,减小了计算量,提高了识别效率,并可以在一定程度上提高识别精度。目前本文只讨论了该结合方法在单纯的动态手势识别中的应用,而相对于孤立词识别来说,连续动态手势的识别难度要大得多,例如中国手语,需要在所有模型构成的状态图中找到一条通过各网格点的路径,这有待更深入的研究与应用。

参考文献

- [1] 孙晓明,陈一民,陈养彬,等.基于视觉的动态手势识别[J].计算机工程与设计,2005,26(11):2934-2936.
- [2] 吴江琴,高文,宋益波.HMM 方法在手语识别中的应用[J].计算机工程与科学,1999,21(3):40-44.
- [3] 姚桂林,姚鸿勋,姜峰.一种基于 DTW-ISODATA 算法的多层分类器手语识别方法[J].计算机工程与应用,2005,8.
- [4] 冯小虎.动态规划思想在算法设计中的应用[J].安徽电子信息职业技术学院学报,2004,3(2).
- [5] Schlenzig J, Hunter E, Jain R. Recursive Identification of Gesture Inputs Using Hidden Markov Models[C]//Proc. 2nd Annual Conf on Applications of Computer Vision. 1994:187-194.
- [6] Campbell L W, Becker D A, Azarbayejani A, et al. Invariant Features for 3-D Gesture Recognition[C]//Proc. 2nd International Workshop on Face and Gesture Recognition, Killington VT, 1996.
- [7] Lee C, Xu Y. Online, Interactive Learning of Gestures for Human Robot Interfaces[J]. IEEE Int Conf on Robotics and Automation, 1996,3(1):30-42.
- [8] Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989,77(2):267-296.