计算机科学1999 Vol. 26№. 10

基于 ANN/HMM 的手语识别方法*)

Sign Language Recognition Method Based on ANN/HMM

吴江琴　高　文

(哈尔滨工业大学计算机科学系　哈尔滨150001)

Abstract Sign Language is the language used by deaf-mute. In this paper the ANN/HMM-based sign language recognition method is proposed. Feature-mapper on posture, position and orientation is built respectively using ANN. The feature vector composed of output from each feature-mapper is used as input of HMM. In addition, in the process of building feature-mapper on posture, multi-feature multi-classifier fusion algorithm is presented. It is proved by experiment that ANN/HMM-based sign language recognition method is feasible and effective.

Keywords sign language recognition, Artificial neural network, Hidden markov model, Feature-mapper

1　引言

手语是聋哑人使用的语言,它是由手形动作辅之以表情姿势而构成的比较稳定的表达系统,是一种靠动作/视觉进行交际的特殊语言。手语识别系统与手语合成系统,共同构成"人-机手语翻译器",为聋哑人提供更好的服务。

人类交互往往声情并茂,除了采用自然语言(口语、书面语言)外,人体语言(表情、体势、手势)也是人类交互的基本方式之一,与人类交互相比,人机交互还呆板得多,因而研究人体语言理解,即人体语言的感知,及人体语言与自然语言的信息融合对于提高计算机的人类语言理解水平和加强人机接口的可实用性是极有意义的[1]。作为人体语言的一个重要组成部分,手语包含信息量最多,它与语音及书面语等自然语言的表达能力相同,因而在人机交互方面,手语完全可以作为一种手段。手语识别作为人体语言理解的重要部分不但具有深远的研究意义,而且具有广阔的实际应用前景。

从识别的技术上来看[2],主要采用:基于人工神经网络(ANN)的识别方法及基于隐 Markov 模型(HMM)的方法。神经网方法具有分类特性及抗干扰性,但其处理时间序列的能力不强,目前主要用于静态手势的识别。一般结构下的 HMM 方法能够有效地处理手势信号的时间特性,因而在手语识别领域一直占有主导地位,如卡内基·梅隆大学的美国手语识别系统[3]及台湾大学的台湾手语识别系统[4]等均采用 HMM 作为识别方法。然而 HMM 拓扑结构的一般性致使该模型在分析手语信号时过于复杂。特别是对于连续的或半连续的 HMM,需要计算大量的状态概率密度和需要估计大量的参数,因而一般手语识别系统均采用离散的 HMM,对于标准的 HMM,它的一个主要局限在于要求对应于每个状态手势段的手势向量是独立的,并且缺少分类特性。因此,这里我们给出了神经网络与 HMM 的混合方法作为手语的训练识别方法,以增加识别方法的分类特性和减少模型的估计参数的个数。我们的实验结果表明将 ANN-HMM 混合方法应用于有18个传感器的 CyberGlove 型号数据手套的中国手语识别系统中是有效和可行的。

2　ANN/HMM 混合模型

由手及手臂的模型可知,每一只手及手臂有27个自由度,两只手及手臂有54个自由度,假设每一个自由度只有两个(实际是大于2个)输出(0,弯曲;1,伸展),

*)本研究得到国家863计划(合同号:863-306-03-01-1)和国家自然科学基金重点课题(批准号:69789301)的资助,还得到国家教委跨世纪人才基金和中国科学院百人计划的资助。吴江琴　博士研究生,研究领域为模式识别与多媒体技术。高文　教授,博士生导师,研究领域为智能计算机接口与多媒体技术。

则手语状态空间中的状态数即为2^{54}，而中国手语的整个手语集只不过3300多个词，可见手语状态空间冗余极大。这里，我们利用 ANN 建立特征映射器，分别将刻画手语信号的各个特征的信息映射为一维标号，以降低手语状态空间的维数。另外，一般结构下的 HMM 方法能够有效地处理手势信号的时间特性，因此我们选取 HMM 作为中国手语的识别方法。然而对于连续的或半连续的 HMM，需要计算大量的状态概率密度并需要估计大量的参数，因此为了提高识别方法的速度，我们采用离散的 HMM。各 ANN 特征映射器的输出构成的特征向量，作为 HMM 的输入。

2.1 ANN 特征映射器

在利用特征映射器进行特征映射时我们采用了 BP 学习算法来建造各个映射器。使用该方法的优势在于：网络参数一经确定，基于训练样本的表示及决策就确定了。另外，该方法可解决识别空间冗余问题，具有抗干扰性和对原始传感数据进行分类的能力。

针对 BP 神经网络收敛速度太慢的缺点，我们采用了 BP 网的快速学习算法——单参数动态搜索算法[5]，其学习速度有明显提高。

2.1.1 *单参数动态搜索算法* 该学习算法采用单变量轮换搜索方式，即每步搜索只让网络的一个参数变化，进行精确的一维搜索。对网络参数按一定顺序轮流进行搜索就构成了 BP 网络的快速学习算法——单参数动态搜索算法。

根据多层神经网络及误差函数的特点，该算法每步搜索只针对有变化的误差部分进行计算，例如：令训练样本点个数为 K，某一训练输入矢量为 $X_k=(x_{k1},x_{k2},\cdots,x_{kn})^T(k=1,\cdots,K)$，网络实际输出为 $Y_k=(y_{k1},y_{k2},\cdots,y_{km})^T(k=1,\cdots,K)$，对应输入 X_k 的理想输出为 $Y'_k=(y'_{k1},y'_{k2},\cdots,y'_{km})^T$，隐含层单元输入向量为 $S_k=(s_{k1},s_{k2},\cdots,s_{kp})^T$，输出向量为 $B_k=(b_{k1},b_{k2},\cdots,b_{kp})^T$，输出层输入向量 $L_k=(l_{k1},l_{k2},\cdots,l_{km})^T$，输入层至隐含层连接权 $\{\omega_{ij}\}(i=1,\cdots,n;j=1,\cdots,p)$，隐含层至输出层连接权 $\{v_{jk}\}(j=1,\cdots,p;k=1,\cdots,m)$，隐含层各单元输出阈值 $\{\theta_j\}(j=1,\cdots,p)$，输出层各单元输出阈值 $\{\gamma_k\}(k=1,\cdots,m)$，误差函数为 $\sum_{j=1}^{m}\sum_{k=1}^{K}(Y_{kj}-Y_{kj})^2$。其中 n 为输入层单元数，p 为隐含层单元数，m 为输出层单元数。在对输出单元阈值 $\{\gamma_i\}(i=1,\cdots,m)$ 进行调整时，误差函数为 $\tilde{f}=\sum_{k=1}^{K}(Y'_{ki}-Y_{ki})^2$，其中，$Y_{ki}=1/(1+e^{-\lambda L_{ki}})$，$\lambda$ 为固定的正值常数。$L_{ki}=(L_{ki}-\gamma_i)+\gamma_i=W(k)+a$，其中，$W(k)$ 为原输出层该单元的输入。从而有：

$$\tilde{f}(a)=\sum_{k=1}^{K}\left(\frac{1}{(1+e^{-\lambda(W(k)+a)}}-Y_{ki}\right)^2$$

求出 $\hat{a}$ 使 $\tilde{f}(\hat{a})=\min\tilde{f}(a)$，$\hat{a}$ 即是 γ_i 的新值，同时调整 $L_{ki},Y_{ki},k=1,\cdots,K$。

该算法尽管增加了一维搜索的次数，但大大减少了误差函数的计算量，进而加快了训练的收敛速度。

同理可对隐层到输出层的连接权值 $\{v_{ji}\}(i=1,\cdots,p;j=1,\cdots,m)$、隐含层阈值 $\{\theta_i\}(i=1,\cdots,p)$、输入层至隐含层连接权 $\{\omega_{ij}(i=1,\cdots,n;j=1,\cdots,p)$ 进行调整。

2.1.2 *多特征多分类器融合算法* 在利用特征映射器进行手形特征映射过程中引入了多特征多分类器融合算法，大大提高了手形特征映射的准确率，从而提高了手语词的识别率。首先根据基本手形的特点，将表征基本手形的特征集分为若干类（可相交）。然后对于同一训练样本集针对每一特征子集进行分类器训练，得到相应的分类器参数集。最后对于通过手势序列分割及手形特征提取的每一有效帧，根据各分类器对应的参数集进行融合特征映射。具体数学描述如下：

设特征集 M 的基数为 $|M|$，模式分类数为 P，将用以分类的特征集 M 分成 k 个可相交的特征子集 M_i，显然 $|M_i|\leqslant|M|,i=1,\cdots,k$，且 $\bigcup_{i=1}^{k}M_i=M$。

对于每一特征子集 $M_l,l=1,2,\cdots,k$，利用单参数动态搜索算法学习对应于该特征子集的分类器 C_l 相应的 BP 网参数 $\{\omega_{ij}\}(i=1,2,\cdots,|M_l|;j=1,2,\cdots,H),\{v_{jk}\}(j=1,2,\cdots,H;k=1,\cdots,p),\{\theta_j^{(l)}\}(j=1,2,\cdots,p;l=1,\cdots,k),\{\gamma_k^{(l)}\}(k=1,\cdots,P;l=1,\cdots,k)$。其中 $\{\omega_{ij}\}$ 为输入层至隐含层连接权，$\{v_{jk}\}$ 为隐含层至输出层连接权，$\{\theta_j\}$ 为隐含层各单元输出阈值，$\{\gamma_k\}$ 为输出层各单元输出阈值，$|M_i|,i=1,\cdots,k$ 为各分类器输入层单元数即特征数，H 为隐含层单元数，P 为输出层单元数即模式类数。

对于任一数据流，利用手势序列分割模块得到有效帧序列为 $X(1),X(2),\cdots X(T)$，对于每一帧数据 $X(t)=(x_1(t)x_2(t)\cdots x_{|M|}(t)),t=1,\cdots T$，记其基于特征子集 $M_l,l=1,\cdots,k$ 的数据为 $X^{(l)}(t)=(x_1^{(l)}(t),\cdots,x_{|M_l|}^{(l)}(t))$，计算其相对于分类器 C_l 的实际输出与第 q 个模式类（$q=1,\cdots,P$）理想输出的误差 $E_q^{(l)}(t)$：

$$E_q^{(l)}(t)=\sum_{j=1}^{P}\left(\left(f\left(\sum_{i=1}^{H}v_{ij}^{(l)}\left(f\left(\sum_{i=1}^{|M_l|}\omega_{ij}^{(l)}x_i^{(l)}(t)\right)+\theta_i^{(l)}\right)\right)+\gamma_j^{(l)}\right)-T_q\right)^2$$

其中 $T_q=(t_{q1},t_{q2},\cdots,t_{qm})^T$ 为对应第 q 个模式类的理想输出，f 为 Sigmoid 函数。

对给定误差精度 ε，定义 $X(t)$ 隶属于第 q 个模式类的隶属度 $A_q(X(t))$ 为：

$$A_q(X(t))=|\{l:E_q^{(l)}(t)\leqslant\varepsilon\}|/P$$

其中$|\cdot|$表示集合的基数。$q_0(t)=\arg\max_q A_q(X(t))$即为$X(t)$所属模式类，$t=1,\cdots,T$。

2.2 HMM学习与识别方法

人的手势可表示成双随机过程，它是内部状态不能直接观察到的马尔柯夫过程。对于这样过程中的每一状态转移，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以精确、定义好的方式确定。用HMM来模型化人的手势，使得处理手势行为的高度随机性成为可能。

对于连续的或半连续的HMM，由于需要计算大量的状态概率密度及需要估计大量的参数，因而我们的手语识别系统所采用的HMM为离散的且每一手语词手势对应一个HMM模型。

一个离散的HMM通常简单描述为$\lambda=\{A,B,\pi\}$，其中$A=\{a_{ij}\}$为状态转移，$B=\{b_{jk}\}$为j状态时观察符号v_k具有标号k的概率，$\pi=\{\pi_i\}$为初始状态分布。

HMM有三个主要问题：评估、估计及解码问题。我们直接关心的是评估及估计问题。评估问题用于解决手势识别问题。估计问题用来产生用于手势识别的各个手势的HMM。

评估问题即手势识别问题，即采集数据与模型的吻合问题。假定识别系统的词汇表共包括V个词条，首先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$O=O_1,O_2,\cdots O_T$；然后针对每个手语词模型$\lambda_l=(A_l,B_l,\pi_l)$计算观察序列的概率$P\{O|\lambda_l\}$$(1\leqslant l\leqslant V)$；最后计算产生序列$L=\arg\max_{1\leqslant l\leqslant V}[P(O|\lambda_l)]$。这里采用了比较有效的"向前-向后"法。

参数估计：给定训练手势集来优化模型参数$\lambda=(A,B,\pi)$使得$P(O|\lambda)$是局部最大的，引用了Baum-Welch算法。Baum-Welch算法是按照最大似然准则来优化模型参数。尽管对于一给定结构的HMM，参数空间有许多局部最优点，但在实际随机实验应用中Baum-Welch算法还是比较有效的。通常HMM的训练采用批处理方式，即所有的手势样本同时作为Baum-Welch算法的输入。这里我们采用了一种不同的方法，首先使用了一个或少数训练样本，运行Baum-Welch算法直至收敛，然后迭代地加入少量样本，更新模型。

3 实验

对照实验室现有的手语合成系统，我们从中国手语字典[7]中选取了120个手语词（见表1）作为系统的实验对象。在我们的系统中，我们选取：右手基本手形为16个，左手基本手形为9个；左右手朝向均为上、下、左、右、前、后等6个方向；右手相对于身体的位置为在头部（包括在额际、在太阳穴部，在耳朵部，在眼部）及在胸部等2个位置，左手相对于身体的位置为在头部及在胸部两个位置；右手的运动轨迹与左手的运动轨迹均为直线、圆周及弧形轨迹。

由于本实验是针对模拟数据的，未考虑运动轨迹特征，即未实现近似样条匹配模块。

如图1所示，首先，利用最大分词法将语句级的输入文本分成词。

然后，根据系统的输入设备数据手套及跟踪器的输出参数（见图2(a)）与合成系统中刻画手及手臂变化的参数（见图2(b)）间的对应关系，在手语合成系统手势库中对应于每个手语词的合成数据基础上加入20%的高斯噪声，得到对应于每个手语词的模拟数据，并在手语词间随机加上噪声帧，存放在系统基本数据库中。引用Quam等[8]的大多数用户手语操作错误率为10%的结论，表明利用该方式生成的模拟数据进行识别是足够强健的。

接着，利用手势序列分割模块对语句级训练样本进行分词，得到对应于每个手语词的训练样本，并利用特征提取模块提取手形、位置及方向特征，通过下面训练过程，得出对应于每个手语词的HMM模型。

训练过程：利用单参数动态搜索算法对训练样本进行训练，训练出对应于右手基本手形的多特征多分类器、对应于左手基本手形的多特征分类器、对应于右手基本位置、右手基本朝向、左手基本位置及左手基本朝向所对应的特征映射器的参数集。利用该参数集将每个手语词的有效帧序列分别进行编码，得到编码序列，作为HMM的输入。最后利用Baum-Welch算法得到对应于每个手语词的HMM模型。

同样，利用手势序列分割及特征提取模块对每个测试样本进行分词及特征提取，得到对应于每个手语词的关于特征空间的测试样本，通过下列识别过程进行词语级识别，得到候选语意。

识别过程：对于每一手语词测试样本，利用单参数动态搜索算法对其每一有效帧进行编码，得到一观察序列，再利用HMM的向前-向后算法选取具有最大概率的HMM模型，其所对应的手语词意，即为该子测试样本的候选语意。

最后，对照语法库，利用二元文法进行校准输出测试样本语意。

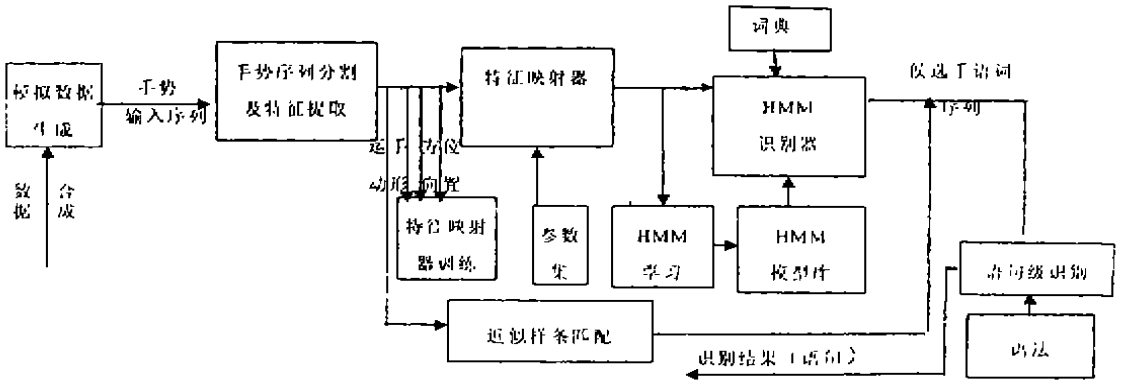
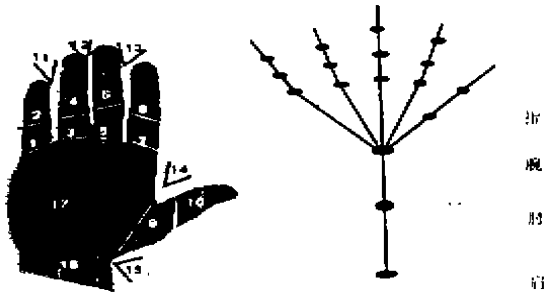


图1　系统总体结构图

实验结果表明，孤立词识别率为90%，简单语句级识别率为92%。



(a) Cyberglove 测量角度示意图 (b) 手及手臂关节模型

图2　手的物理模型及 Cyberglove 测量角度示意图

参 考 文 献

1 Gao W. Enhanced User Interface by Hand Gesture Recognition, Chinese J. of Advanced Software Research (Allerton Press Inc., New York), 1996, 3(1): 30～42

2 Watson R. A Survey of Gesture Recognition Techniques: [technical report TCD-CD-93-11]. Department of Computer Science, Trinity College, Dublin 2, 1993

3 Starner T., Pentland A. Real-time American sign language recognition from video using hidden Markov models, MIT media Lab Perceptual Computing Section, TR-375, 1996

4 Liang R., Ouhyoung M. A Sign language recognition system using hidden Markov model and context sensitive search. In: Proc. of the ACM Symposium on VR Software and Technology. Hongkong, 1996. 59～66

5 王雪峰，冯英俊. 一种新的神经网络学算法. 哈尔滨工业大学学报，1997. 1

6 Rabiner L R, Juang H. An Introduction to Hidden Markov Models. IEEE ASSP Magazine, 1986 (Jan.): 4～16

7 中国聋人协会. 中国手语. 北京华夏出版社，1991

8 Quam D, et al. An Experimental Determination of Human Hand Accuracy with a Dataglove. In: Proc. Human. Factors Society 33rd Annual Meeting, Vol. 1. 1989. 315～319

表1　词汇集

识别词汇集	我；你；我们；自己；大家；谁；同志；明天；后天；晚上；中午；白天；上午；下午；以前；将来；去；到；看；听；说；参观；需要；谢谢；注意；对不起；再见；希望；满意；相信；信任；承认；否定；支持；发言；报告；论文；能力；发明；常常；可能；一定；为；有；不；是；对；能；最；向；哈尔滨；要；社会；残疾人；盲人；同事；同学；朋友；校长；教师；教授；导师；专家；知识分子；学生；星期一；星期二；星期三；星期四；星期五；星期六；星期日；一星期；今天；进；工作；休息；上班；下班；加班；开会；开始；欢迎；联系；介绍；帮助；完成；请；按时；尊敬；关心；放心；信心；耐心；研究；准备；会议；参加；决定；总结；毕业；内容；学习；上课；下课；读书；教育；教学；学校；中学；专业；放学；考试；手语；严格；技术；科研；科学；重要；掌握。