基于隐马尔科夫模型的中国手语识别

Chinese sign language recognition based on Hidden Markov Model

信息学院 电子工程与信息科学系

崔 骁 PB09210297

李厚强 教授

二〇一三年六月

中国科学技术大学

University of Science and Technology of China

本科毕业论文

A Dissertation for the Bachelor's Degree

基于隐马尔科夫模型的中国手语识别

Chinese sign language recognition based on Hidden Markov Model

姓　　名	崔 骁
B.S. Candidate	Xiao Cui
导　　师	李厚强 教授
Supervisor	Prof.Houqiang Li

二〇一三年六月

June,2013

中国科学技术大学

学士学位论文

题　　目　基于隐马尔科夫模型的中国手语识别

院　　系　信息学院 电子工程与信息科学系

姓　　名　崔 骁

学　　号　PB09210297

导　　师　李厚强 教授

二〇一三年六月

University of Science and Technology of China

A Dissertation for the Bachelor's Degree

Chinese sign language recognition based on Hidden Markov Model

B.S. Candidate Xiao Cui

Supervisor Prof.Houqiang Li

Hefei, Anhui 230026, China

June,2013

致　　谢

四年的大学时光接近尾声，似一坛多年窖藏的好酒，还未来得及细细品尝，便已只剩得杯中残酒，那感觉也和酒后一般，如梦如幻，如痴如醉，大学四年，美好而灿烂。

对于我来说，在大四这一年加入手语组并在这里完成毕设，不但对科研有了基本的了解，学到了许多做科研的方法，同时也让我对科研有了新的看法和理解，收获颇丰。

我想感谢我的指导老师李厚强教授。李老师一直非常关心手语组的科研情况，经常在百忙之中抽出时间和我们进行讨论，李老师渊博的学识，严谨的作风和他对科研的执着都令我获益匪浅。

感谢指导我的师兄唐傲，他旺盛的精力，强大的工作能力和开阔的眼界带给我另一片天地，更让我对优秀一词加深了认识，此外，唐傲师兄在工作上的要求和指导也给了我非常大的帮助。

我还要感谢手语组的成员们，是你们在组会上的讨论让我不断改进，是你们在实践中的建议让我不断提高。

最后感谢所有关心、支持和鼓励我的亲人和朋友们。尤其感谢我的家人一直以来对我的支持与关怀。谢谢你们！

目　录

摘　要

连续手语识别可以分为三个步骤：连续手语的分割、手语孤立词的识别和自然语言处理的过程。连续手语识别是一个复杂的系统，因为连续手语中没有明确的断点，并且其视觉识别过程非常复杂，连续手语识别比与其类似的语音识别要困难得多。

本文首先介绍了作者利用隐马尔科夫模型实现手语孤立词汇识别的过程。然后分别讨论了精简循环网络和条件随机场在连续手语分割中的应用。针对每个模型，本文首先介绍数学背景和模型的基本结构，然后介绍实现中的相应细节，最后介绍实验结果和结论。

本文使用中国科大多媒体计算与通信教育部微软联合实验室利用 Kinect 采集的基于视觉的数据进行识别，孤立词的识别实验效果良好，连续手语的分割也取得了积极的实验结果。

关键字:　手语孤立词识别，隐马尔科夫模型，连续手语分割，精简循环网络，条件随机场

Abstract

Continuous sign language recognition includes three procedures: continuous sign language segmentation, isolated words recognition and the process of natural language processing. Continuous sign language recognition is a complex system. It is much more difficult to recognize continuous sign language than speech which is similar to sign language, because there' s no definite segmentation point in continuous sign language, and that the recognition process through vision itself is difficult.

The paper first introduces the isolated sign words recognition process based on Hidden Markov Model (HMM). Next, the paper discusses the implementation of Simple Recurrent Network (SRN) and Conditional Random Field (CRF). For each model, we first introduce mathematical ideas and the basic structures. Then, we introduce details of implementation and finally, we introduce experiment results and conclusions.

The data used in the paper which is captured by Kinect is provided by China Ministry of Education–Microsoft Key Laboratory of Multimedia Computing and Communication, University of Science and Technology of China. Based on the data, we get good results in both the experiments of isolated sign words recognition and continuous sign language segmentation.

Keywords: isolated sign words recognition, HMM, continuous sign language segmentation, SRN, CRF

第一章　绪论

第一节　本文研究的背景和意义

据统计，全世界有 5 亿左右的聋哑人，中国就有 2700 万人。除了使用文字，在日常生活中聋哑人主要通过手语同他人进行交流沟通。然而，由于手语的普及程度较差，聋哑人与正常人之间的沟通存在困难, 即使是聋哑人之间，也会因为地域差异而难以交流，例如中国手语和美国手语，甚至在中国大陆的不同地区，差异也是明显的存在的，正如口语中的普通话和粤语的区别。在这样的情况下，通过计算机来自动的识别聋哑人的手语，再将手语翻译成文本或者语音，将具有重要的现实意义。

人机交互技术的研究是计算机技术研究领域的重要组成部分。当人与人进行面对面的通讯时, 包括口语及书面语等自然语言与包括手语、表情及口型等人体语言传递信息。因而研究人体语言的感知模型及其与自然语言的信息融合, 对于提高计算机自然语言理解水平和加强人机接口的可实用性是极有意义的。手语识别作为人体语言理解的一部分, 有着非常重要的作用。一方面, 它是虚拟现实人机交互的主要手段; 另一方面它又是聋哑人利用计算机与正常人交流的辅助工具。每个手语是由一个手势序列组成, 而每个手势是由手形变化序列组成。根据手语输入介质的不同, 手语识别系统可分为两种: 基于视觉的手语识别系统和基于设备输入 (如数据手套、位置跟踪器等) 的手语识别系统。本文所处理的数据是基于视觉获取的。

第二节　国内外研究历史现状

目前，大多数的手语识别都是基于小词汇集特定人操作的孤立手语词识别，然而手语是固定的大词汇集、语法约束完备的手势集，以手语识别为开端，

可以为将来研究普遍的手势分析工作积累经验。手语既包括手部信息，如手形、朝向、位置等，还包括身体其他部分的运动，例如：人脸表情、头势、躯干运动等，手语识别需要同时处理这些并行信息，对它们进行同步、融合。手语的高度结构性使得它可以作为一个很好的研究平台，国内外都有不少学者正在进行深入的研究。

台湾大学的 Liang 等 [1] 专注台湾手语识别的研究，其实现实时连续手语识别的研究基于数据手套所采集的数据，利用姿势、位置、方向和动作等特征参数，51 个自定义的基础姿势，8 个方向的 HMM 模型，实现基于自定义的 250 个词汇的识别。

麻省理工学院的 Starner 等 [2] 在 1995 年利用 HMM 模型实现了一个基于单目视觉的美国手语识别系统，该系统能够识别由 40 个美国手语词组成的简单连续语句。

北京大学高文等人 [3] 提出一种将连续手语识别分解成各孤立词识别的分治方法，用于非特定人连续手语识别。利用精简循环网络实现连续手语的分割，而后利用 HMM 模型进行孤立词的识别，达到了比较好的效果。

Sagawa 等人 [4] 利用 Cyber-glove 获得手语操作者两只手上各关节实时的位置信息计算出手型、手的速度等变化信息来对连续手语的分割点做出判断。而后对手语词和动增分量做出判断。并对每个手语词建立简单的属性 (attributes) 来进行手语词汇的识别。

C. Vogler 等 [5] 则提出了对手语词之间过渡动作部分的处理方法，即去前一个词的结束点和后一个词的开始点分别聚簇到 8 点，通过做搜索方法来完成手语词的分割。而后使用 HMM 模型进行孤立词的识别。

Helen Cooper 等 [6] 提出的给手语词建立 sub-units 的方法很好的解决了手语的语义模型的建模问题，同时也很好的提高了孤立词的识别率。只是对于连续手语的识别则未做许多研究。

总的来说，大部分的识别都是基于 HMM 模型或者 HMM 模型的改进做出的，而连续手语的分割则成为连续手语识别的一个瓶颈，虽然有不少分割方法，总的来说得到的好效果对环境和数据的要求都太高。

第三节　本文研究工作概述

本文使用 HMM 模型对简单的 12 维空间位置特征条件下的手语孤立词进行识别。本论文尝试将精简循环网络和条件随机场模型运用到手语识别中，对序列数据进行分割，进而实现连续手语中孤立词的分解，为识别连续的手语句子做好前期的准备，并对二者的分割效果做了简单的比较。

第四节　本文的组织结构

第 2 章主要介绍了隐马尔科夫模型的基本定义、三个基本问题的求解。第 3 章为其在孤立词识别当中的应用。第 4 章介绍了精简循环网的基本模型和其在连续手语分割中的应用。第 5 章介绍了条件随机场的定义。第 6 章为其在连续手语分割中的应用。第七章为全文总结和未来展望。

第二章　HMM 模型

本章主要介绍了隐马尔科夫模型 (Hidden Markov Model) 的基本理论。同时对该模型的三个典型问题进行了探讨与求解，三个典型问题即评估问题、解码问题和学习问题。

第一节　HMM 定义

在描述隐马尔可夫模型之前，首先介绍 Markov 链的定义。Markov 链是 Markov 随机过程的特殊情况，即 Markov 链是状态和时间都离散的 Markov 过程。数学上给出如下定义：

随机序列 Xn，在任一时刻 n，它可以处在状态 $S_1, S_2, S_3...S_N$，且它在 $m+k$ 时刻所在的状态为 q_{m+k} 的概率只与它在 m 时刻的状态有关，而在 m 时刻以前所处的状态无关，即：

$$\begin{aligned} &P(X_{m+k} = q_{m+k} | X_m = q_m, X_{m-1} = q_{m-1}, ..., X_1 = q_1) \\ &= P(X_{m+k} = q_{m+k} | X_m = q_m) \end{aligned} \tag{2-1}$$

式中，$q_m \epsilon (S_1, S_2, ...S_N)$。则称为 Markov 链，并且称

$$P_{ij}(m, m+k) = P(q_{m+k} = S_i | q_m = S_j) \tag{2-2}$$

（其中 $1 \leq i, j \leq N, m, k$ 为正整数）为 k 步转移概率，当它与 m 无关时，称此 Markov 链为齐次 Markov 链，此时

$$P_{ij}(m, m+k) = P_{ij}(k) \tag{2-3}$$

当 k=1 时，称为一步转移概率，简称转移概率，记为 a_{ij}，所有转移概率可以构成一个转移概率矩阵：

$$A = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \dots & a_{NN} \end{pmatrix} \qquad 0 \le a_{ij} \le 1, \sum_{j=1}^{N} a_{ij} = 1, 1 \le i \le N \tag{2-4}$$

除了转移矩阵，还需要初始化，由此引进概率矢量 $\pi = (\pi_1, \pi_2 ... \pi_N)$，其中

$$\pi_i = P(q_i = S_i) \qquad 1 \le i \le N, 0 \le \pi_i \le 1, \sum_{i=1}^{N} \pi = 1 \tag{2-5}$$

至此，我们已经知道了什么是 Markov 链。接下去我们给出隐马尔科夫模型中的五个概率参数定义 [7]：

1. N：模型中 Markov 链的状态数。各个状态表示为 $S = \{S_1, S_2, ..., S_N\}$。记 t 时刻的马尔科夫链所处的状态为 q_t，显然 $q_t \epsilon (S_1, S_2, ..., S_N)$。

2. M：每个状态对应的可能观察值的数目。各个观察值可记为 $V = \{v_1, v_2, ..., v_M\}$。记 t 时刻的观察值为 O_t，显然 $O_t \epsilon (V_1, V_2, ..., V_M)$。

3. A：状态转移概率矩阵。$A = a_{ij}$，其中

$$a_{ij} = P\left[q_{t+1} = S_j | q_t = S_i\right], \qquad 1 \le i, j \le N. \tag{2-6}$$

   即为 t 时刻状态 i 在 t+1 时刻转移到状态 j 的概率。

4. B：观察序列的概率矩阵。$B = b_j(k)$，其中

$$b_j(k) = P\left[O_j = v_k | q_t = S_j\right], \qquad \begin{matrix} 1 \le j \le N \\ 1 \le k \le M. \end{matrix} \tag{2-7}$$

   即为在 q_j 状态时，观察序列符号为 v_k 的概率。

5. π ：初始状态概率的集合。$\pi = \pi_i$，其中

$$\pi_i = P\left[q_1 = S_i\right], \qquad 1 \le i \le N. \tag{2-8}$$

所以，我们可以将隐马尔科夫模型记为：$\lambda = (N, M, \pi, A, B)$。该模型一般简记为 $\lambda = (\pi, A, B)$。

第二节 HMM 的三个基本问题

[7] 在上一节给出的隐马尔科夫模型的前提之下，我们需要解决三个基本问题，以使模型能够更好地得到应用。

一、 HMM 的评估问题

评估问题是指在已知观察序列 $O = O_1O_2 \cdots O_T$ 和模型 $\lambda = (A, B, \pi)$ 的情况下，如何有效的计算出该模型下观察序列发生的概率 $P(O|\lambda)$。

通常，我们采用“前向 -后向”算法来计算该概率。

定义前向变量如下：

$$\alpha_t(i) = P(O_1, O_2, ..., O_t, q_t = i|\lambda) \tag{2-9}$$

表示当已知模型参数 λ 和观察序列 $O = O_1, O_2, ..., O_{t-1}$ 出现的概率，观察向量 O_t 在第 i 个状态节点出现的概率。

初始化得：

$$\alpha_1(i) = \pi_1 b_1(O_1) \qquad 1 \le i \le N \tag{2-10}$$

通过迭代可得：

$$\alpha_{t+1}(j) = \sum_i \alpha_t(i) a_{ij} b_j(O_{t+1}) \qquad 1 \le t \le T-1, 1 \le j \le N \tag{2-11}$$

所以，

$$P(O|\lambda) = \sum_i \alpha_T(i) \tag{2-12}$$

至此，评估问题已经得到解决。同理，我们还可以定义后向变量如下：

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, ..., O_T, q_t = i|\lambda) \tag{2-13}$$

表示当已知模型参数 λ 和观察序列 $O_{t+1}, O_{t+2} \cdots O_T$ 出现的概率，观察向量 O_t 在第 i 个状态结点出现的概率。

初始化得：

$$\beta_T(i) = 1 \qquad 1 \le i \le N \tag{2-14}$$

通过迭代得：

$$\beta_t(j) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \qquad t = T-1, T-2, ...1, 1 \le i \le N \tag{2-15}$$

所以，

$$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(O_1) \beta_1(i) \tag{2-16}$$

如果将两向量结合求解，则为：

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \beta_t(i) \qquad 1 \le t \le T \tag{2-17}$$

前向向量和后向向量还将用于学习问题的解决中。

二、 HMM 的解码问题

解码问题是在给定观察序列 $O = O_1, O_2 \dots O_T$ 和已知模型参数 $\lambda = (\pi, A, B)$ 的条件下，求解出最佳状态序列：

$$Q^* = q_1^*, q_2^*, \dots, q_T^* \tag{2-18}$$

这里的最佳序列是指使得 $P(O|\lambda)$ 最大的状态序列。

为了求解该问题，需要定义：

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} p(q_1, q_2, \cdots, q_t = S_i, O_1, O_2, \cdots, O_t|\lambda) \tag{2-19}$$

表示 t 时刻沿一条路径 $q_1, q_2, \dots, q_t$ 且 $q_t = S_i$，产生 $O_1, O_2, \dots, O_t$ 的最大概率。同时定义 $\psi_t(i)$ 表示 t 时刻状态为 S_i 条件下的前一个状态。则，最佳状态求解过程如下：

初始化：

$$\delta_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N \tag{2-20}$$

$$\psi_1(i) = 0 \qquad 1 \le i \le N \tag{2-21}$$

通过递归有：

$$\delta_t(j) = \max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]\,b_j(O_t) \qquad 2 \le t \le T, 1 \le j \le N \tag{2-22}$$

$$\psi_t(j) = \arg\max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}] \qquad 2 \le t \le T, 1 \le j \le N \tag{2-23}$$

终止条件为：

$$P^* = \max_{i\le i\le N}[\delta_T(i)] \tag{2-24}$$

$$q_T^* = \arg\max_{1\le i\le N}[\delta_t(i)] \tag{2-25}$$

状态序列求取为：

$$q_T^* = \psi_{t+1}(q_{t+1}^*) \qquad t = T-1, T-2, \ldots, 1 \tag{2-26}$$

至此，最佳序列问题就得到了解决。

三、 HMM 的学习问题

隐马尔科夫模型的学习问题，即 HMM 参数估计问题，也就是在给定观察序列 $O = O_1, O_2, \cdots, O_T$，如何确定一个模型参数 $\lambda = (\pi, A, B)$ 使得 $P(O|\lambda)$ 最大。参数学习的主要步骤如下：

1. 模型初始化。即初始化模型的各部分参数。一般采取均匀处理。
2. 利用训练样本通过评估问题，计算出前向和后向的概率，并重估所需参数。
3. 通过重估公式计算得到一组新的参数。
4. 计算重估后的新参数是否收敛。

上述模型的学习步骤，其实就是 Baum-Welch 算法，也就是模型的训练过程，通过该过程最终可以得到较为理想的模型参数，进而为之后的识别工作做准备。

本文所谈论的为观察序列为离散值的情形。定义：

$$\varepsilon_t(i,j) = P(q_t = i, q_{t+1} = j|O, \lambda) = \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)} \tag{2-27}$$

表示给定观察序列和模型时，该模型 t 时刻处于状态 i，t+1 时刻处于状态 j 的概率。

那么 t 时刻处于模型 i 的概率就是对 t+1 时刻所有概率的求和，如下：

$$\varepsilon_t(i) = \sum_{j=1}^{N} \varepsilon_t(i,j) \tag{2-28}$$

而利用我们前面介绍的前向和后向概率，我们可以求得：

$$\varepsilon_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \tag{2-29}$$

于是我们可以有如下重估公式：

$$\overline{\pi_1} = \varepsilon_1(i) \tag{2-30}$$

$$\overline{a_{ij}} = \frac{\text{从状态}S_i\text{过渡到状态}S_j\text{的平均次数}}{\text{从状态}S_i\text{向其它状态转移的平均次数}} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i,j)}{\sum_{t=1}^{T-1} \varepsilon_t(i)} \tag{2-31}$$

$$\overline{b_{ij}} = \frac{\text{处于状态}S_j\text{和出现观察 k 的平均次数}}{\text{处于状态}S_j\text{的平均次数}} = \frac{\sum_{t=1,O_k=V_k}^{T} \varepsilon_t(i,j)}{\sum_{t=1}^{T} \varepsilon_t(i)} \tag{2-32}$$

至此，隐马尔科夫模型的学习问题也得到了解决。也就是在给定训练集的条件下，我们可以产生出该条件下的最佳模型。

第三节　本章小结

本章主要介绍了隐马尔科夫模型的基本参数，基本框架，以及该模型的三个典型问题的求解。其中利用前向后向算法求解评估问题，即已知模型参数 λ 和观察序列 $O = O_1, O_2, \cdots, O_T$，能够有效的计算该模型下观察序列发生的概率；利用维特比算法求解解码问题，即给定观察序列和已知模型参数 λ 的条件下，求解出最佳状态序列；利用 Baum-Welch 算法求解学习问题，即隐马尔科夫模型的参数估计问题，也就是给定观察序列，确定一个新的模型参数使得 $P(O|\lambda)$ 最大。

第三章　隐马尔科夫模型识别孤立词

通过前面的介绍，我们已经知道什么是 HMM 模型，以及如何求解其三个典型的问题。本章主要介绍如何将 HMM 模型运用在手语识别的数据处理中。利用 HMM 学习问题的求解，可以将一部分样本进行训练从而获得模型的参数。利用 HMM 评估问题的求解，可以将剩余的识别样本进行概率求解，求得的最大概率所对应的模型，即为识别的结果。本章还给出了在不同码书下的实验结果，并进行了对比。

第一节　HMM 中手语特征简要描述

本文的重心不在特征的提取上，而是对特征提取后的后期识别进行主要的实验研究，所以本文只对手语特征提取与处理进行简要的描述。中国科学技术大学多媒体计算与通信教育部 -微软重点实验室的手语小组已经利用 Kinect 设备对手语视频进行特征提取，在提取的数据中包含多维特征，包括轨迹追踪、人体骨骼框架、深度信息以及彩色信息等等。在描述手语的特征获取过程中，有十二维向量特征表示一帧数据信息和一个状态特征表示一帧数据信息。由于考虑到 HMM 模型本身的特性，所以采用状态特征作为 HMM 模型输入的主要特征，进而实现模型参数的学习和识别。

在利用之前提到每帧状态特征时，状态是用 0 到 99 这 100 个状态表示，也就是说如果我们直接将其使用到 HMM 模型中的话，模型的观察序列元素的可能取值应该为 100，我们称之为码书。但，实验表明当码书过大时，会造成数据过于稀疏，不仅收敛速度慢，而且造成很大的计算负担，更糟糕的是，码书到一定程度时，其并不是和正确率成正比。所以我们对 100 个状态进行矢量量化，量化参数可以手动选择，本文进行了三种量化，分别是 25、50 和 75 码书量化。再对其进行训练以及识别。

本模型所使用的词汇为“上”，“下”，“运动”，“衣服”，“校长”，“小孩”，“洗脸”，“东”，“西”，“南”，“北”，“篮球”，“高兴”，“搬运”，“翱翔”，“游泳”等共 18 个孤立词汇，每个词汇有十五个样本，样本是由五个人中每人交叉三次进行测试提取，并非是基于特定人的数据，使得样本相对而言更具有代表性。

第二节　手语词 HMM 模型的建立

从上述 HMM 的学习问题的求解中，我们可以利用手语数据每个孤立词的前十个样本进行训练，从而建立每个词汇的 HMM 模型。建立模型的流程图如下3.1 所示：

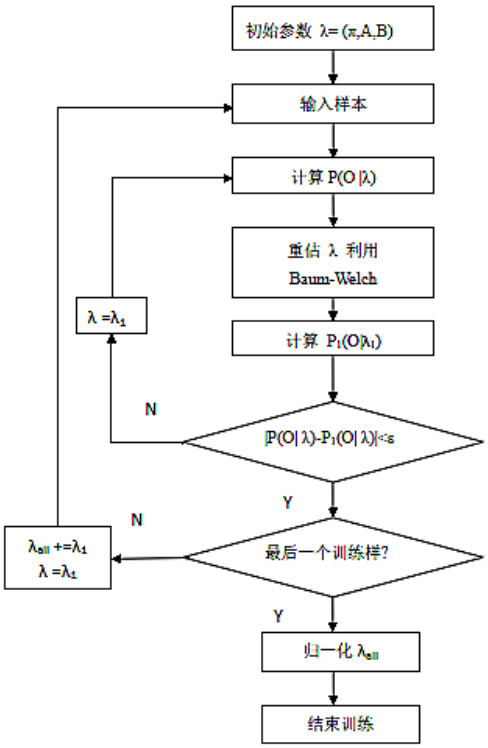

图 3.1 HMM 模型训练流程图

第三节　手语词的识别

如上所述，我们已经得到了 18 个孤立词的 HMM 模型参数，则我们可以利用 HMM 模型中的评估问题对样本剩余的五个样本进行评估，得到的最大概率即为识别的结果，识别过程流程图如下3.2所示：

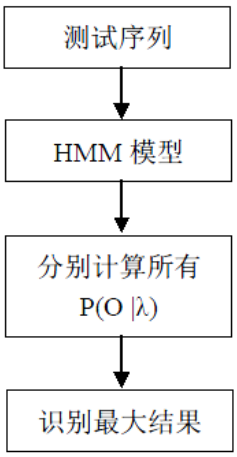

图 3.2 识别过程

第四节　实验结果与分析

表3.1给出了十个样本训练，其余五个样本识别的结果，表中 n 表示模型的隐含状态数，m 为码书数。表中“3/5”表示五个识别样本中识别正确的个数为三，以此类推。

从表3.1中数据可知，随着码书的递增，识别率并不是与码书成正比关系，在状态相等情况下，码书为 50 时可以获得较高的准确率，并且为 50 时，实验的计算量相对比较均衡。说明码书的大小与识别结果呈现先上升后下降的关系。

由表3.1可知，该模型在识别“小孩”和“下”时出现严重的偏差，原因可能是该两个词的特征提取时没有提取好，导致训练样本和识别样本之间的差距太大，所以不能正确的识别，解决的办法可以通过增加训练样本数，让模型得到充分的学习。其中 n 为每个词所包含的状态数，经试验，n=6 时效果最好。

	n=6;m=25;	n=6;m=50;	n=6;m=75;	单个识别率
左	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	60%
运动	$\frac{4}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	93%
右	$\frac{4}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	93%
游泳	$\frac{3}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	87%
衣服	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	100%
校长	$\frac{2}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	67%
小孩	$\frac{1}{5}$	$\frac{0}{5}$	$\frac{0}{5}$	7%
下	$\frac{0}{5}$	$\frac{0}{5}$	$\frac{0}{5}$	0
洗脸	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	80%
西	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	60%
上	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{4}{5}$	93%
南	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	100%
篮球	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	100%
高兴	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	100%
东	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{3}{5}$	67%
北	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	60%
搬运	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	47%
翱翔	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	80%
总识别率	67%	76%	70%	

表 3.1 18 个手语词的识别结果

第五节 本章小结

本章主要介绍如何利用 HMM 模型对手语数据进行训练以及识别。本文采用 18 个孤立词的十五个样本，前十个样本用于训练 HMM 模型，生成每个词对应的 HMM 模型参数。训练所使用的方法是 HMM 模型的参数学习算法，前一章已经讨论过其求解方法。在识别过程中，对剩下的五个样本分别输进训练好的模型，利用前一章介绍的评估问题可以求得最大概率的模型，也就是识别的结果。通过实验结果，我们可以发现在码书等于 50 的时候可以获得较好的识别率。至此，就实现了手语的孤立词的识别。

第四章　基于 SRN 的边界检测

第一节　SRN 及其改进

Elman 在循环网基础上进行了改进，提出了精简循环网 (SRN)[8]。循环网通过反馈层的引入而使网络具备了记忆和利用上文的能力，并已经成功地应用到语音识别、手写体识别和孤立手语词识别等方面。典型的精简循环网络共有 4 层神经元。令网络在接受第 t 时刻输入向量 I_t 时相应各层的输出为：隐层 H_t，反馈层 C_t，输出层 O_t，W_C^H, W_I^H, W_H^O 分别表示反馈层到隐层、输入层到隐层、隐层到输出层的权值矩阵，见图4.1。

反馈层节点是隐层节点的拷贝，并加入了一个单位的延迟。网络的输入层和反馈层组成联合输入层。若 Φ, Ψ 分别为隐藏层神经元和输出层神经元的阈值向量，则有

$$H_t = f\left(C_t \cdot W_C^H + I_t \cdot W_t^H - \Phi\right) \tag{4-1}$$

$$O_t = f\left(H_t \cdot W_H^O - \Psi\right) \tag{4-2}$$

神经元的活跃函数 $f(\cdot)$ 一般取为 Sigmoid 函数。精简循环网络采用误差反向传播 (back-propagation) 算法 [9] 来进行网络参数训练。

反馈层的引入，使得网络输出不但与当前的网络输入有关，而且与前一时刻的网络状态有关，而前一时刻的网络状态又是前面所有输入计算的结果。因此，精简循环网络可以记忆并使用较大范围的上文。

精简循环网络通过隐藏层，利用所有前导输入的信息作出当前判断，但是无法利用下文，为此，这里进行改进。将下文矢量作为当前输入的一部分，从而使上下文的微结构信息同时得到有效利用，这样，新的输入为 $I_t = [I_t I_{t+1}]$，其他计算同 SRN。

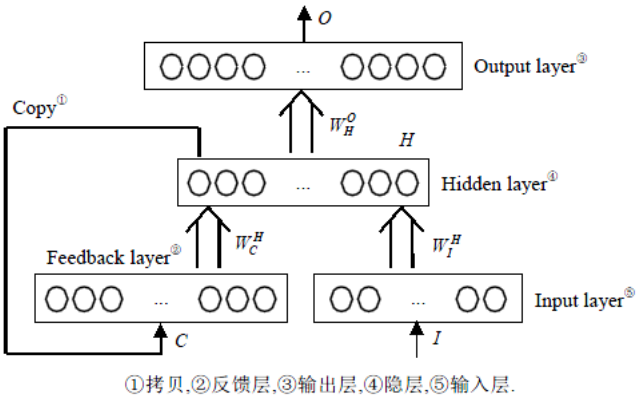

图 4.1 精简循环网络

第二节 SRN 检测手语边界

SRN 训练流程 [3][8][10]：

输入经过标记的句子序列，经过标记是指给句子打上标签，例如：我爱中国科技大学，经过标记为 sil, I, sil, love, sil, China, sil, technology, sil, university, sil。这一步骤是事先由人工来完成。

正如上一章所说，手语数据是利用 Kinect 采集 12 维的位置信息经 KDtree 之后分类为 0-99 共 100 个状态，然后对每一帧的状态进行二进制编码，需 7 位，同时引进下文单元 7 位作为输入，这样共选用 14 个输入节点，节点的输入值为 $I_i^t \epsilon \{0,1\}, i = 1,2,\ldots,14.$

由于连续手语中间没有停顿，不能确定其边界，这里采用自动分割的方法。输出层共有 6 个节点。取 4 帧长度的窗，当前两帧属于前一个 label 而后两帧属于 sil 的时候，窗中的第一帧标为 [0,1,0,0,0,0]（非 sil 类的 label 中），第二帧标为 [0,0,1,0,0,0]（非 sil 类的右边界），第三帧为 [0,0,0,1,0,0] （sil 类的左边界），第四帧标为 [0,0,0,0,1,0]（sil 类的中间）；当前两帧属于 sil 而后两帧属于另一 label 的时候，窗中的第一帧为 [0,0,0,0,1,0]，第二帧为 [0,0,0,0,0,1]，第三帧为 [1,0,0,0,0,0]，第四帧为 [0,1,0,0,0,0]。

上面的分割是根据联合概率最大原则，其中的概率使用该帧状态在整个库中属于某个词的概率来统计。Sil 类的状态是取库中每个词的前面三帧和后面三帧的状态做统计的一部分，其中，每个词的样本为 15 个；另一部分则取最初人工分割时分割点前后各 3 帧（人工分割时在词与词之间只标记分割点，未分

割出 sil 类的状态）。而一般词的 label 则对每个词的库来统计。最后有 100*21 个概率（100 个状态，20 个词（label）+sil）。然后再通过孤立词模型来判断每段是哪个词。

SRN 的训练模型中引入自适应学习速率、附加动量项的反向传播算法作为 SRN 的基本学习算法，将状态文件作为训练集，进行二进制编码，与下文一起输入到 SRN 网中，计算出相应各时刻的输出，将该输出与理想输出的误差反向传播，根据误差调整 SRN 的网络权值。学习开始时，权值矩阵、隐层神经元的阈值都赋予 (-1,+1) 区间内的随机值，反馈单元初始化为 0.5. 学习速率初始值为 0.01，附加动量项 0.95，期望误差最小值为 0.01，最大循环次数为 200000 次。

第三节　实验结果与分析

取 12 个“大家好”视频进行训练，剩余 3 个进行测试。其中每段分割点数最大为 12，得到如表的结果，表中未 0 的地方表示不分段。

“大家好”	人工标记分割点	srn 求出的分割点					
测试数据 1	44	0	48	0	0	0	0
	49	0	54	0	0	0	0
测试数据 2	55	0	57	98	0	0	0
	60	0	63	105	0	0	0
测试数据 3	56	56	79	0	98	0	0
	61	62	96	0	115	0	0

表 4.1 srn 求出分割点与人工标记分割点

由表4.1可知，对于“大家好”句子的分割，SRN 模型能达到较高的识别率，与人工标记的分割点相比，人工标记的标准分割点 SRN 都能识别出来，只是 SRN 模型找出了更多的分割点，这与训练数据的分布和获取的方式即训练中目标输出的确定有关。同时前期的特征提取的提升对于手语句子的分割也有重大的作用。

第四节　本章小结

本章主要介绍了经过改进的 SRN 模型的基本结构和本文作者实验所采用的训练过程。经过不断实验改进得到的本章的训练流程，实现了对“大家好”句子的分割，达到了预期的效果。有关效果的提升则寄望于与模型的改进和数据质量的提升。

第五章　CRF 模型

条件随机场 (Conditional Random Fields，CRF) 是由 Lafferty 等人 [11] 于 2001 年提出。它可以看成是一个无向图模型或马尔可夫随机场，是一种用来标记和切分序列化数据的统计框架模型。目前，条件随机场被应用于解决分词、新词识别、命名实体识别等自然语言处理任务，取得了良好的效果。本章主要介绍 CRF 基本框架以及主要原理 [12][13]。

第一节　CRF 的图结构

条件随机场是一种条件模型，它不需要 HMM 所要求的严格的独立假设，并且不是有向图模型，而是无向图模型。CRF 条件随机场是在给定观测序列的条件下定义的关于整个类别标记的一个单一的联合概率分布，而不是在给定当前状态的条件下，定义下一个状态的状态分布。这里标记序列的条件概率取决于观测序列的非独立、相互作用的特征。

CRF 条件随机场模型是无向图模型的一种形式，在给定将要标记的观测序列的情况下，无向图模型可以被用来在标记序列上定义一个联合概率分布。假设 X，Y 分别表示需要标记的观察序列和它对用的标记序列的联合分布随机变量，条件随机场 (X，Y) 就是一个以观测序列 X 为全局条件的无向图模型。

通常我们定义一个无向图 $G = (V, E)$，其中 V 表示结点，E 表示边，则我们的标记序列就对应于每一个结点，即有 $Y = Y_v|v\epsilon V$。整个图的分布以观察寻列 X 为条件，则与 G 相关联的概率为 $P(y_1, y_2, \cdots, y_n|X)$。如果每个随机变量 Y_v 满足关于 G 的马尔科夫属性，给定 X, Y_v 以外的所有随机变量，有以下：

$$P(Y_v|X, Y_u, u \neq v) = P(Y_v|X, Y_u, u \sim v) \tag{5-1}$$

其中 $u \sim v$ 表示两者在图 G 中相邻，那么 (X, Y) 就是一个条件随机场。

理论上该图的结构可以是任意的，但是用于标记任务建模时，最通用的是一阶链式（First-order Chain）。

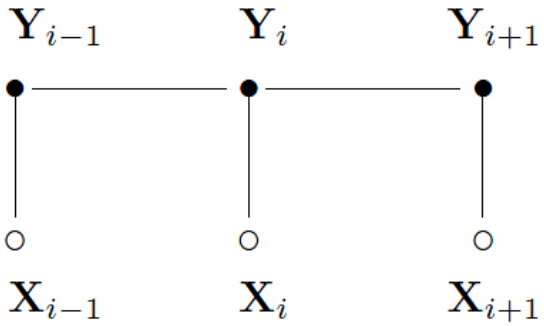

图 5.1 条件随机场的一阶链式结构

如图5.1所示，我们给出了条件随机场模型的基本图结构，而且在图中，我们可以注意到 X 只是我们的观察序列，在条件随机场模型中 X 只是当作一种条件而已，因此我们并不需要对 X 做任何的独立性假设。

第二节 CRF 的势函数表示

因为 CRF 的模型是一种无向图模型，所以该模型的结构可以用一个归一化的势函数的乘积来表示。乘积中的每个因子都是正值实函数。由无向图中相关定义知，如果 G 中两个顶点之间没有边，则意味着两个顶点表示的随机变量独立于 G 中其它给定的顶点。所以如果两个结点如果在物理上没有直接相连的话，它们不会出现在同一个势函数中。解决这种问题的方法是利用最大的全通环 (clique)。每个最大的全通环有且仅有一个势函数表示，那么这就确保了势函数所涉及的任何随机变量对，其顶点是直接联系的。

尽管无向图模型中随机变量的联合分布可写成势函数的乘积，需要指出的是一个孤立的势函数并没有直接的概率意义，而是表示了定义这个势函数所涉及的随机变量的结构上的约束而已。这反过来也影响了全局结构的概率，即一个概率大的全局结构较概率小的全局结构更能满足这些约束条件。在给定观测序列 X 的情况下，Lafferty 等定义了标记序列 Y 的概率是势函数乘积的一个归

一化形式，其中每个因子形式如式5-2:

$$\exp\left(\sum_j \lambda_j t_j(Y_{i-1}, Y_i, X, i) + \sum_k \mu_k s_k(Y_i, X, i)\right) \tag{5-2}$$

这里 $t_j(Y_{i-1}, Y_i, X, i)$ 是关于整个观察序列和位置 i 以及位置 $i-1$ 的标记的特征函数，$s_k(Y_i, X, i)$ 是关于整个观察序列和位置 i 的状态特征函数。其中系数是模型的参数，是可以在训练过程中估计得到的。

为了方便表示，我们可以将特征函数统一表示成如下形式：

$$f_j(Y_{i-1}, Y_i, X, i) \tag{5-3}$$

则当给定一个观测序列 $X = X_1, X_2, \cdots, X_i, \cdots, X_n$，对应的标记序列 $Y = Y_1, Y_2, \cdots, Y_i, \cdots, Y_n$ 的概率如下所示：

$$P(Y|X, \lambda) = \frac{1}{Z(X)} \exp\left(\sum_j \sum_i \lambda_j f_j(Y_{j-1}, Y_j, X, i)\right) \tag{5-4}$$

其中 $Z(X)$ 是归一化函因子，形式如下所示：

$$Z(X) = \sum_Y \exp\left(\sum_j \sum_i \lambda_j f_j(Y_{j-1}, Y_j, X, i)\right) \tag{5-5}$$

至此，$P(Y|X)$ 就由势函数表示出来了。

第三节　CRF 的参数估计

CRF 的主要任务之一就是从训练数据中估计出特征参数，可以采用多种方法，本文采用的是最大似然估计法。设 Λ 为参 λ 的某一集合，而 T 为训练集，则：

$$\begin{aligned} L_\Lambda &= \sum_T \log P(Y|X, \lambda) \\ &= \sum_T \log \frac{1}{Z(X)} \exp(\sum_j \sum_i \lambda_j f_j(y_{i-1}, y_i, X, i)) \\ &= \sum_T \left(\sum_j \sum_i \lambda_j f_j(y_{i-1}, y_i, X, i) - \log Z(X)\right) \end{aligned} \tag{5-6}$$

$$\Lambda^* = \arg\max_{\lambda} \sum_{T} \log P(Y|X,\lambda) \tag{5-7}$$

其中 Λ^* 即是估计出来的最佳参数。

由于 L_Λ 为凸函数，导数为零的点记为最值点，则可对其进行求偏导：

$$\frac{\partial L_\Lambda}{\partial \lambda_j} = \sum_{T} \left(\sum_{j} \sum_{i} \lambda_j f_j(y_{i-1}, y_i, X, i) - E_{P(Y|X)[F_K(Y,X)]} \right) \tag{5-8}$$

$$\frac{\partial L_\Lambda}{\partial \lambda_j} = O_j - E_j = 0 \tag{5-9}$$

其中 $E_P(Y|X)[F_K(Y,X)]$ 为第 K 个训练序列的特征期望，O_j 为 λ_j 在 T 中出现的频率，而：

$$E_j = \sum_{T} E_P(Y|X)[F_k(Y,X)] \tag{5-10}$$

是 λ_j 在模型分布中的特征期望。直接对其计算代价是很大的，之后我们将会介绍动态规划的方法对其求解。

在参数估计的过程中，如若直接使用最大似然估计法，可能会导致过度学习的问题，我们可以通过引入惩罚项 $\frac{\sum_j \lambda_j^2}{2\sigma^2}$，则：

$$L_{\Lambda'} = L_\Lambda - \frac{\sum_j \lambda_j^2}{2\sigma^2} + const \tag{5-11}$$

则其导数为：

$$\frac{\partial L_{\Lambda'}}{\partial \lambda^j} = \frac{\partial L_\Lambda}{\partial \lambda_j} - \frac{\lambda_j}{\sigma^2} \tag{5-12}$$

至此，参数估计问题就可以运用最优化方法进行解答了。一般可以使用 GIS、IIS 等迭代法。本文使用的最优化方法是 MIT 提供的 L_BFGS 算法。

第四节　CRF 的矩阵计算

对于链式结构的 CRF 标记中，我们可以认为的多添加两个标记以表示开始和结束标记，分别用 Y_0 和 Y_{n+1} 表示。同时设 Ψ 为标记集合，标记的个数用 L 表示，则我们可以定义 $n+1$ 个矩阵 $M_j(X)|i = 1, \cdots, n+1$，其中每个矩阵都是一个 $L * L$ 维的矩阵，每个元素为如下：

$$M_i(y', y|X) = \exp\left(\sum_{j} \lambda_j f_j(y', y, X, i) \right) \tag{5-13}$$

则当给定观察序列 X 的时候我们就可以用矩阵形式表示概率：

$$P(Y|X,\lambda) = \frac{1}{Z(X)}\Pi_{i=1}^{n+1}M_i(y_{i-1},y_i|X) \tag{5-14}$$

其中，归一化因子为：

$$Z(X) = \left[\Pi_{i=1}^{n+1}M_i(X)\right]_{start,end} \tag{5-15}$$

式中，start 代表 Y_0，end 代表 Y_{n+1}。则 CRF 的矩阵描述就可以给出了：

$$P(Y|X,\lambda) = \frac{\Pi_{i=1}^{n+1}M_i(y_{i-1},y_1|X)}{\left[\Pi_{i=1}^{n+1}M_i(X)\right]_{start,end}} \tag{5-16}$$

第五节 CRF 的动态规划

在对 CRF 模型的参数进行估计时可以选择多种算法，但无论采取什么样的训练算法都需要有效的计算每个特征函数在模型分布下对于训练集上每个观察序列 X 的期望值，即我们需要求得：

$$\begin{aligned} &E_P(Y|X,\lambda)[F_k(Y,X)] \\ &= \frac{1}{Z(X)}\sum_{Y}\left[\exp\left(\sum_j\sum_i\lambda_j f_j(y_{i-1},y_i|X,i)\right) * \sum_j f_j(y_{i-1},y_i|X,i)\right] \end{aligned} \tag{5-17}$$

如果直接对式进行计算，若 X 有 n 个元素，则 Y 的可能取值就会有 n_L 种，可见计算的花销是巨大的。因此，通常使用类似 HMM 中的前向 -后向算法解决这个问题。所以要定义前向向量 $\alpha_i(x)$ 和后向向量 $\beta_i(X)$ 分别如下：

$$\alpha_0(y|X) = \{^{1 \quad if\ y=y_0}_{0 \quad others} \tag{5-18}$$

$$\beta_{n+1}(y|X) = \{^{1 \quad if\ y=y_0}_{0 \quad others} \tag{5-19}$$

其递归形式如下所示：

$$\alpha_i^T(X) = \alpha_{i-1}^T(X)M_i(X) \tag{5-20}$$

$$\beta_i(X) = M_{i+1}(X)\beta_{i+1}(X) \tag{5-21}$$

因此，

$$E_P(Y|X,\lambda)\left[F_k(Y,X)\right] = \frac{1}{Z(X)}\sum_{i=1}^{n}\alpha_{i-1}(y|X)M_i(y,y'|X)\beta_i(y'|X) \tag{5-22}$$

$$Z(X) = \beta_0(X) = \alpha_{n+1}(X) \tag{5-23}$$

这样我们就可以很快的求出所需要的期望，进而能得到模型的特征参数 λ。

第六节　标记的给出

通过参数估计，学习出模型的参数后，可以利用以下式子进行标记的给出：

$$\begin{aligned} Y^* &= \arg\max_Y P(Y|X) \\ &= \arg\max_Y \frac{1}{Z(X)} \exp\left(\sum_j \sum_i \lambda_j f_j(y_{i-1}, y_i, X, i)\right) \end{aligned} \tag{5-24}$$

利用类似于 HMM 中的维特比算法即可以给出最佳序列的标记。

第七节　本章小结

本章主要介绍了 CRF 条件随机场模型的基本概念，以及相应的模型计算的推导。CRF 是一种无向图，可以利用势函数表示；在评估参数的时候可以用最大似然法通过 L-BFGS 算法计算；在计算特征期望时可以利用动态规划的方法降低计算量，最终实现模型的建立；最后，利用建立好的模型对输入序列给出最佳的标记序列结果。

第六章　CRF 在连续语句中的应用

CRF 模型是一个较新的模型，2001 年才被提出，而且目前尚未被运用到连续手语的识别当中，其主要是运用在文本词汇的分割，例如汉语介词、命名实体等分割，也有最基本的连续序列分段 [14]。该章主要介绍利用 MIT 的工具包实现对手语每帧数据进行标记，进而在标记跳跃点给出分割点。并且根据手语本身特性对结果进行分析。

第一节　CRF 中手语特征描述

前一章节已经对手语特征进行了简要的描述，在 CRF 模型的特征数据输入中，本文使用的是由 Kinect 获取的含深度信息的数据，并对其进行处理获得十二维数据，该十二维数据是对测试者的几个特定位置进行归一化得来，所以不会因为测试者的高低胖瘦而影响手语的数据获取。

之所以使用十二维的数据是因为该数据相比于一维的状态数据更丰富，并且注意到 CRF 模型的参数时和数据的维度是相关的，而 CRF 模型参数太少不能得到好的识别结果。在主要实验之余，作者曾利用一维状态数据作为输入，给出的分段结果为 0%，由于不在主要实验讨论范围内，所以此处不对其做详细讨论。本章主要讨论十二维数据的识别效果。

第二节　CRF 模型训练与识别

CRF 模型训练的输入为 18 个词的前十个样本并依此标记从 0 到 17，数据输入为 CSV 文件，以符合 MIT 工具包。将 MIT 工具包修改成分离训练与识别，并且符合十二维数据模板的新工具包。将其输入到修改过的 MIT 工具包

中，通过 L-BFGS 算法进行最优参数的求解，就可以生成输出模型文件供识别所用。

在识别过程中，利用训练得到的模型，通过维特比算法给出最终的标记序列。

由于我们对每个词给定一个标记，也就是说在训练的过程中我们训练的转移参数是当前标记到当前标记的转移，这一部分是稀疏的，由于目前我们训练的数据不是连续的句子，所以不同标记之间的转移参数并没有得到学习，将保持初始权重。而状态特征函数训练的参数是当前数据的本质特征，在当前标记转移到当前标记的权重不能起到主导作用，而当前的本质特征起主要作用的时候，就会发生标记之间的转移。

得出最佳的状态标记序列之后如何判别分割点又是另一部分工作，理论上不同标记之间的转移应该就是分割点，但通过实验我们可以得出多种标记之间的转移但非分割点的情况。我们将连续 15 帧以上相同标记认为是有意义的标记，在连续十五以上的标记末尾认为是分割点。训练与识别的框图可如6.1所示：

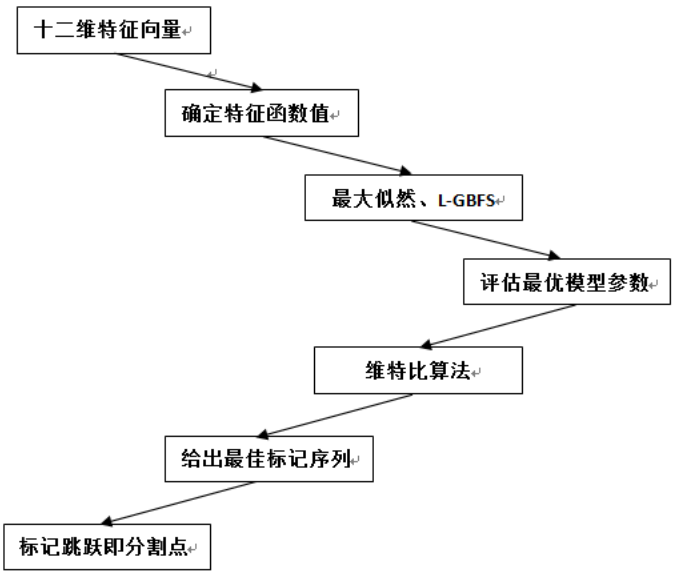

图 6.1 crf 模型训练与识别框图

第三节　实验结果与分析

通过上述的训练以及输入需要给出标记的样本序列后，就可以得到序列标记，对序列标记进行处理可以得到分割点。如下两表所示，它们都是对“大家好”两个词组成的句子的十五个样本进行分割，该样本并非由训练样本组成。我们将前后的噪声称为“silence”，其中表6.1是包含 silence 部分的分割结果，而表6.2是去除 silence 后的分割结果。

表中给出的分割点，我们有理由认为在动态浮动的 10 帧范围内为正确有效的分割，也就是在三分之一秒以内给出分割点都是认为是正确的，因为在该短时间内的分割基本不会影响孤立词的识别。由此我们可以统计出两表的分割正确率。表中的 0 表示该次实验未能给出分割点，即将“大家”和“好”两个词识别成一个词。

样本	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
正确分割点	48	68	56	41	40	47	46	36	32	49	50	49	44	55	56
实验分割点	49	37	24	30	41	40	41	38	31	16	21	16	31	53	42

表 6.1 “大家好”含 silence 的分割结果

分割正确率：50

样本	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
正确分割点	34	34	37	33	30	35	37	36	32	35	31	35	33	39	31
实验分割点	34	0	32	25	30	27	31	37	30	0	0	0	19	36	16

表 6.2 “大家好”去除 silence 的分割结果

分割正确率：60

对“大家好”的分割结果可以得知，分割效果并不是很好，但在去除 silence 后正确率有了明显的提升，说明周围环境以及噪声会对实验结果造成较大的影响；而在不能分割的样本中，返回看样本视频，发现测试者的速度过快，模型将其认为是某一个动作，未能识别出跳转区间，导致不能给出分割点；造成正确率不高的原因还在于手语本身的特性，因为相同的动作可能属于不同的词，所以给模型的识别带来了较大的歧义和难度。

对于噪声的影响，可以在提取数据时选择特定的友好环境；对于不能分割情况，可以要求测试者适当降低自己手语的比划速度，让 Kinect 能够充分提取出手语特征；而对于手语本身相关性，我们无法改变手语本身，但是我们可以加强参数的训练，例如可以使用连续句子进行训练，或者是直接提取句子中词的转移部分进行训练，以寻求更好的模型。

第四节　本章小结

本章简要介绍了用于 CRF 模型的手语数据类型，并且给出了训练模型和识别给出序列标记的方法以及框图。利用“大家好”的样本数据进行了包含噪声和去除噪声的分割实验，分析了造成分割率不高的原因以及给出了今后可能的改进方法。虽然分割效果不能达到百分之百的效果，但是证明了 CRF 在手语连续词中对于分割的一定可行性。将对今后的工作起到重要的引导作用。

第七章　总结和展望

本文首先介绍了利用 HMM 模型实现手语孤立词汇识别的过程。然后着重讨论了 SRN 和 CRF 在连续手语分割中的应用，介绍了实现中的相应细节，实现了比较好的效果。

从实验结果来看，SRN 和 CRF 都实现了找到“大家好”句子中词与词之间分割点的任务，但是由于实验样本较少，它们的实验结果仍缺乏一定的代表性，需要后来实验者进行进一步的探索。

同时，显而易见的是，在实验当中，简单的 12 维输入数据也是制约着分割实验的一大因素。后来的实验者应继续丰富特征来提高系统的精度，作者也从一些文献当中了解到手型特征的加入会对连续手语句子的分割起到不小的提升作用。连续手语句子的分割实际上是一个分类问题，更好的特征则意味着更好的分类效果。

非特定人的连续手语识别是一项前沿技术，这其中连续手语的分割则是制约的一大瓶颈，至今国际上未出现能得到多数认同的效果明显的分割方法，本文作者在这个领域进行了积极有益的探索，然而，限于作者水平有限，无法在有限的毕设期间做出高水平的前沿成果，还寄望于后来者能够突破重重难题，早日完成成熟可用的手语识别系统，造福千万聋哑人！

参考文献

[1] Ouhyoung M. Liang R H. A real-time continuous alphabetic sign language to speech conversion vr system. *Computer Graphics Forum*, pages 67–76, August 1995.

[2] Starner T. Visual recognition of american sign language using hidden markov models. Master's thesis, MIT,Media lab, Feb 1995.

[3] Gao W. Fang G. Zhao D. et al. A chinese sign language recognition system based on sofm/srn/hmm. *Pattern Recognition*, 37(12):2389–2402, 2004.

[4] Masaru Takeuchi Hirohiko Sagawa. A method for recognizing a sequence of sign language words represented in a japanese sign language sentence. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 434–439, 2000.

[5] D. Metaxas C. Vogler. Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. *IEEE International Conference on Systems, Man and Cybernetics*, pages 156–161, 1997.

[6] Helen Cooper Eng-Jon Ong et al. Sign language recognition using sub-units. *Journal of Machine Learning Research*, pages 2205–2231, 2012.

[7] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, pages 257–286, 1989.

[8] Jeffrey L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, pages 195–225, 1991.

[9] Fernando J Pineda. Generalization of back-propagation to recurrent neural networks. *Physical review letters*, 59(19):2229–2232, 1987.

[10] Axel Cleeremans, David Servan-Schreiber, and James L McClelland. Finite state automata and simple recurrent networks. *Neural computation*, 1(3): 372–381, 1989.

[11] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[12] Hanna M Wallach. Conditional random fields: An introduction. *Technical Reports (CIS)*, page 22, 2004.

[13] 韩雪冬，周彩根. 条件随机场综述. 中国科技论文在线.

[14] L-P Morency, Ariadna Quattoni, and Trevor Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[15] Britta B. H. Hienz. Relevant features for video-based continuous sign language recognition. *International Conference on Automatic Face and Gesture Recognition*, pages 440–445, 2000.

[16] Britta Bauer, Hermann Hienz, and K-F Kraiss. Video-based continuous sign language recognition using statistical methods. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 463–466. IEEE, 2000.