基于条件随机域的上下文人类动作识别

朱文球,刘 强
ZHU Wen-qiu,LIU Qiang

湖南工业大学 计算机与通信学院,湖南 株洲 412008
School of Computer and Communication,Hunan University of Technology,Zhuzhou,Hunan 412008,China

ZHU Wen-qiu,LIU Qiang.Conditional Random Fields with loop and its inference algorithm.Computer Engineering and Applications,2008,44(28):180-183.

Abstract: A new algorithm for human motion recognition based on Conditional Random Fields (CRFs) and Hidden Markov Models(HMM)—HMCRF is proposed.Most existing approaches to humman motion recognition with hidden states employ a Hidden Markov Model or suitable variant to model motion streams;a significant limitation of these models is the requirement of conditional independence of observations.In contrast,conditional models like the CRFs seamlessly represent contextual dependencies,support efficient,exact inference using dynamic programming,and their parameters can be trained using convex optimization.We introduce conditional graphical models as complementary tools for human motion recognition and present an extensive set,experiments show that the proposed approach outperforms the linear-chain structure CRF and HMM in terms of recognition rates.

Key words: Conditional Random Fields(CRFs);Hidden Markov Models(HMM);junction tree algorthms;human motion recognition

摘 要:提出一种新的基于条件随机域和隐马尔可夫模型(HMM)的人类动作识别方法——HMCRF。目前已有的动作识别方法均使用隐马尔可夫模型及其变型,这些模型一个最突出的不足就是要求观察值相互独立。条件模型很容易表示上下文相关性,且可使用动态规划做到有效且精确的推论,它的参数可以通过凸函数优化训练得到。把条件图形模型应用于动作识别之上,并通过大量的实验表明,所提出的方法在识别正确率方面明显优于一般线性结构的 CRF 和 HMM。

关键词:条件随机域;隐马尔可夫模型;联合树算法;动作识别
DOI:10.3778/j.issn.1002-8331.2008.28.060 **文章编号:**1002-8331(2008)28-0180-04 **文献标识码:**A **中图分类号:**TP391

1 引言

人类动作识别是计算机视觉和模式识别领域的一个研究热门，跟踪和识别一段视频中人的活动有很多潜在的应用,包括智能人机交互、娱乐、安保、自动监视系统。人类动作识别的一个挑战是缺乏一个清楚的分类架构,同样的动作可能经常被同时分类为不同的动作类别。如某些动作可能会同时包含类似的基本动作单元(如跑步和手臂摆动,走与握手),甚至简单动作之间的转换在时间上也会发生意义含糊的部分和重叠现象。目前人类动作识别通常使用的方法是隐马尔可夫模型[1](HMM)或其变形[2],而用这种生成式模型的架构来进行训练与识别通常必须假定在时间序列上动作彼此独立,但事实上相似的动作经常发生在不同的时间点上且通常存在着长距离的相关性,所以单独使用当前的观察样本以及前一个状态可能很难去识别这个时间点上的动作种类。如果在时间点上能同时考虑过去、未来与当前观察值的相邻性,则可降低动作间的模糊程度。条件随机域(CRFs)[3]可以用来避免样本必须独立的假设,但是在 CRFs 相关研究中,往往使用线性链状结构来描述变量之间的相关性,事实上在随机域中的变量相关性可能是非常复杂且具有回路的网状图形模式,尤其是在时间顺序上前后相互影响的变量,这时用精确的树状推论方法便没有办法来正确推论所有变量的联合概率。本文以条件随机域为出发点,将网状图形模型通过转换构造相应的联合树,用联合树的方法来解决 CRFs 中具有回路的复杂情形的所有变量的联合概率。

2 相关研究工作

有大量的文献介绍动作分类,在这里介绍几个与研究工作相关的方法。生成式模型如 HMMs 以及它的变形已成功地运用于基于二维和三维观察对象的动作分类，在生成式模型中,为保证推导的正确性,需要作出严格的独立性假设。事实上,大多数序列数据都不能被表示成一系列独立的元素,相似的动作经常发生在不同的时间点上且通常存在着长距离的相依性。

除了生成式模型以外,分辨式模型也被用于解决序列标注问题。在说话和自然语言处理领域，最大熵马尔可夫模型(MEMM)[4]已经用于单词识别、文本分割和信息提取。

基金项目:湖南省教育厅基金项目(the Project of Department of Education of Hunan Province,China under Grant No.07C233)。
作者简介:朱文球(1969-),男,副教授,高级工程师,硕士生导师,主要从事计算机图像处理、模式识别等研究;刘强(1980-),工程师,研究方向为数字图像处理。
收稿日期:2007-11-19 **修回日期:**2008-03-24

CRFs 最早由 Lafferty 等[3]介绍,并且被广泛应用于自然语言处理领域如名词辅助标注问题、命名体识别和信息提取。最近几年 CRFs 在计算机视觉方面的应用研究逐渐活跃起来,C.Sminchisescu 等[5]应用 CRFs 于人类动作分类(如行走、跳跃等),这个模型利用时间点上过去与未来的相邻状态跟观察值之间的上下文窗口(context window),它采用动态规划可以做到有效且精确的推论,能分辨出一些模糊的动作类型(如正常行走与来回走动),并且效果非常好。Kumar 等[6]使用 CRF 模型于图像区域标注工作。Torralba 等在文献[7]中介绍了 BRFs 模型,该模型组合局部和全局图像信息用于上下文目标识别。

3　条件随机域模型

一般来说,人类动作分类的研究最主要还是以隐马尔可夫模型来模型化人类动作是随着时间序列来进行的特性,而用这种生成式模型的架构来进行训练与识别通常必须假定在时间序列上人体动作彼此独立,事实上相似动作经常发生在不同的时间点上且通常存在长距离的相依性,所以单独使用当时的观察状态以及前一个状态可能很难去辨别某一个时间点上的动作种类。

3.1　一般线性链结构的 CRF

条件随机场(Conditional Random Fields,CRFs)是一种新的概率图模型,它具有表达元素长距离依赖性和交叠性特征的能力、能方便地在模型中包含领域知识、且较好地解决了标注偏置问题等优点。

条件随机场是一种用于在给定输入结点值时计算指定输出结点值的条件概率的无向图模型。若 X 是一个值可以被观察的"输入"随机变量集合,S 是一个值能够被模型预测的"输出"随机变量的集合,且这些输出随机变量之间通过指示依赖关系的无向边所连接。让 $C(S,X)$ 表示这个图中的团的集合,CRFs 将输出随机变量值的条件概率定义为与无向图中各个团的势函数(potential function)的乘积成正比:

$$P_A(S|X)=\frac{1}{Z_X}\prod_{c\in C(S,X)}\Phi_c(S_c,X_c) \tag{1}$$

其中,$\Phi_c(S_c,X_c)$表示团 c 的势函数。当图形模型中的各输出结点被连接成一条线性链的特殊情形下,CRFs 假设在各个输出结点之间存在一阶马尔可夫独立性,二阶或更高阶的模型可类似扩展。若让 $X=(x_1,x_2,\cdots,x_M)$表示被观察的输入数据序列,让 $S=(s_1,s_2,\cdots,s_M)$表示一个状态序列。在给定一个输入序列的情况下,线性链的 CRFs 定义状态序列的条件概率为:

$$P_A(S|X)=\frac{1}{Z_X}\exp\left(\sum_{t=1}^{M}\sum_{k=1}^{K}\lambda_k f_k(s_{t-1},s_t,x,t)\right) \tag{2}$$

其中 $f_k(s_{t-1},s_t,x,t)$是一个任意的特征函数,λ_k 是每个特征函数的权值,Z_X 为归一化因子且

$$Z_X=\sum_s\exp\left(\sum_{t=1}^{M}\sum_k\lambda_k f_k(s_{t-1},s_t,x,t)\right)$$

3.2　网状图形 CRF 模型—HMCRF

3.2.1　HMCRF 及其推导

根据不同的应用模式,可以用不同的数学图形模型来描述在条件随机域中变量之间的关系。当描述的变量之间是层次关系时,可使用树状态结构的 TCRF[7],当描述两个不同路径但是有共同时间索引的马尔可夫模型状态与观察值之间的关系时,可用 DCRF[8]架构。对于观察值特征具有任意重叠的 CRF 结构(如图 1 所示),已经很难用树状推论理论来推论出联合条件概率,用联合树算法可有效地、极大地简化推理计算。本文从联合树的观点出发,将具有复杂网状结构的图形模型三角化从而构造出联合树,提出一种相对简单的条件概率推论算法,称之为 HMCRF。

联合树是指这样的一个无向图 $J=(V_J,E_J)$,它上面所定义的势函数是将变量集的每个实例映射成一个数值。用 Φ_X、Φ_S 分别表示信念域的势和分隔集 S 上的势,那么有 $\sum_{X\setminus S}\phi_X=\phi_S$,$P(U)=\prod_i\phi_{x_i}\Big/\prod_j\phi_{s_j}$。将具有时间重叠的 CRF 用图形表示如图 1,去掉状态与状态转移方向则变成一个无向有环图。

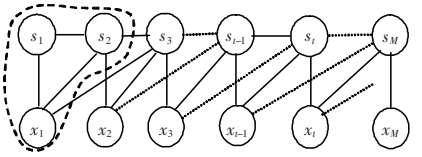

图 1　具有观察值特征重叠的 CRF 架构图
(图中虚线圈起部分为一个子图)

在图 1 所示的架构图中状态之间存在着马尔可夫链的架构,观察值与状态间相互影响的关系因为包含着过去与未来的时间点,所以产生了网状图形架构。标准的树状态图形推论方法便没有办法在这样的架构上实现,但通过分析发现,图 1 符合建构联合树[9]的条件,因此可以依照联合树的步骤来构建一个树状态的结构,从而可以用标准的推论方法来推论这样的图形模型。

为了构造联合树,引入接口(Interface),它表示由时间点 t 中的,在时间点 $t+1$ 中有孩子的结点组成的集合,及转移状态的起始点集合。定义 S 是一个联合树所有的接口的集合,对所有接口具有 $s\in S$,令 $d(s)$表示这个接口所相邻的子图的个数。接口完全分隔时间点,这样就满足时间点上条件独立性的假设。将一个个时间点的联合树粘合在一起就构造出了带回路的 CRF 架构所对应的联合树。

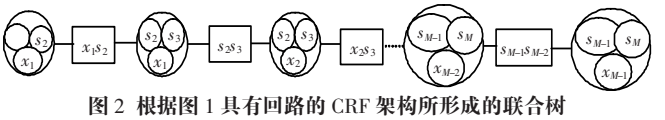

图 2　根据图 1 具有回路的 CRF 架构所形成的联合树
(圆结点代表团结点而方形结点代表接口)

根据上面的分析,图 2 所示联合树的联合概率分布可表示为:

$$P_A(S,X)=\frac{\prod_{c\in C}\phi_c(x_c)}{\prod_{s\in S}[\phi_s(x_s)]^{d(s)-1}}=\prod_{t=1}^{M-2}\frac{\phi(x_t,s_t,s_{t+1})\phi(x_t,s_{t+1},s_{t+2})}{\phi(x_t,s_{t+1})\phi(s_{t+1},s_{t+2})}=\prod_{t=1}^{M-2}\psi(x_t,s_{t+\Delta t}) \tag{3}$$

边界分布 $Z_\theta(X)$可用以下的方法计算:

$$Z_\theta(X)=\sum_s p(X,S)=\sum_s\prod_{t=1}^{M-2}\psi(x_t,s_{t+\Delta t})=\prod_{t=i}\sum_{s_t}\psi(x_t,s_{t+\Delta t}) \tag{4}$$

定义条件概率为:$p(S|X,\theta)=\frac{1}{Z_\theta(X)}\prod_{t=1}^{M-2}\psi(x_t,s_{t+\Delta t})=\frac{1}{Z_\theta(X)}\times$

$\exp(\sum_{k=1}^{B}\lambda_k^{\mathrm{T}}.F_k(S,X))$,其中 $\theta=\{\lambda_1^{\mathrm{T}},\lambda_2^{\mathrm{T}},\cdots,\lambda_B^{\mathrm{T}}\}$为所求模型参数,$B$ 为模型中所定义的所有特征函数的个数。

假设一个完全标注好的训练样本 $X=\{S^d,X^d\}_{d=1..M}$ 是来自于同一概率分布且相互独立,且训练样本的概率分布 $p(S^{(k)}|X^{(k)},\theta)$ 是 θ 的相似度函数,HMCRF 的参数可以通过优化下述对数似然值而求得:

$$L(\theta)=\sum_{d=1}^{M}\log(p_\theta(S^d,X^d))=\sum_{d=1}^{M}[\log\frac{1}{p(X^{(d)})}+\sum_{i=1}^{B}\lambda_i^{\mathrm{T}}.F_i(S^{(d)},X^{(d)})] \quad (5)$$

上述似然函数的最大值可以通过梯度搜索的方法计算求得:

$$\frac{\mathrm{d}L(\theta)}{\mathrm{d}\theta}=\sum_{d=1}^{D}(\sum_{i=1}^{M}\frac{\mathrm{d}_\theta F(S,X^{(i)})}{\mathrm{d}\theta}-\sum_{X}P_\theta(X|S^d)\frac{\mathrm{d}F_\theta(X,S^{(d)})}{\mathrm{d}\theta})=E_{\tilde{p}}[F_\theta(S,X)]-\sum_{X}E_p[F_\theta(S,X^{(i)})] \quad (6)$$

其中 $E_{\tilde{p}}$是训练样本的实验分布,E_p 表示概率分布的期望值。

3.2.2 特征函数定义

在条件式随机域模型中,特征函数的定义在于观察序列与状态序列之间的关系,和状态与状态之间的关系,因此将经过最大化相似度训练的隐马尔可夫模型(HMM)应用在形成特征的过程之中。利用观察值与状态之间的相似度分数,与状态转移概率来定义势函数。

根据以上的描述,针对每一个人类肢体动作中的每一个状态建立一个特征函数为:

$$F_s(S,X)=\varphi_\theta(x_t,S_t)+\varphi_\theta(x_t,x_{t+1},s_t,s_{t+1})=\sum_{i=1}^{M}\delta(s_t=s).p(x_t|s_t) \quad (7)$$

$$\varphi_\theta(x_t,S_t)=\sum_{j=1}^{A}\lambda_j f_j(x_t,S)$$

$$\varphi_\theta(x_t,x_{t+1},s_t,s_{t+1})=\sum_{b=1}^{B}\beta_b g_b(x_t,x_{t+1},s_t,s_{t+1})$$

$p(x_t|s_t)$代表相似度。若假设在条件随机域模型中势函数所代表的函数值是正数,每一个肢体运动序列对于不同的肢体动作模型,就可以求出对应特征函数的值,当取得这些特征函数的值之后,就可以建构出条件随机域模型,如式(3)所示。

3.2.3 参数的训练

训练的目标是为了最大化条件条件概率 $P(c|X)$,其中 c 为类别数,X 为观察样本序列。由式(5)、(6)延伸,在每一个类别 c 中都存在一个特定的状态序列,换句话说,某一种特定的状态序列可以表示某一个类别。因此在应用中,希望最大化 $P(c|X)$等同于最大化 $p(S\in c|X)$,由式(6)知道,当一个特征函数的实验期望值与真实模型期望值相等时,会使得对数相似度为最大。

在实验中,每一个接受训练的模型都有其相对应的训练样本,由 HMM 训练中的结果,可以得到这些样本对自己所属模型的相似度值,因此可以求出每一个相对应的特征函数值,最后再将相同的特征函数在训练样本中得到的分数求和,就得到特征函数的实验期望值:

$$E_{\tilde{p}}=\sum_{X\in c_i}F_s(S,X)=\sum_{X\in c_i}\sum_{t}\delta(s_t=s)p(x_t|s_t) \quad (8)$$

在计算特征函数的真实模型期望值时,要考虑每一个观察序列对每一个类别都会有影响,因此不能认为观察序列所属的类别为已知。计算方式为:

$$E_p=\sum_{\forall X}\sum_{\forall c}P(c|X)F_s(S,X) \quad (9)$$

在计算某个特征函数 E_{λ_i} 时,这个状态只出现在某个类别中,因此在计算 $\sum_{\forall c}P(c|X)$时只会有一个被加进来。例如,当要计算 $F_{s\in c_1}(S,X)$这个特征函数的真实模型时,则是去求

$$E_{s\in c_1}=\sum_{\forall X}P(c_1|X)F_{s\in c_1}(S,X) \quad (10)$$

其中 $P(c_1|X)$的部分,利用条件式随机域的辨识部分就可以求得,而模型参数的部分则是利用上一个训练回合所得的模型参数带入,$F_{s\in c_1}(S,X)$则由 HMM 训练中的结果可得,因此可计算出每一个特征函数的真实模型期望值。模型参数的训练与更新可以按照梯度下降算法求得:

(1)根据式子(8),计算每一个特征函数的实验期望值 $E_{\tilde{p}}$

(2)初始化参数 a,阈值 θ,$\eta(.)$,令 $k=0$,$a(0)=1$

(3)do

$k=k+1$;

根据第 k 次循环所更新的参数,计算每一个特征函数的真实模型期望值 E_p;

$$a_{(k+1)}=a_{(k)}-\eta_{(k)}*\log(\frac{E_p}{E_{\tilde{p}}})$$

Until $|\log(\frac{E_p}{E_{\tilde{p}}})|<\theta$。

3.2.4 动作的识别与时间复杂度计算

根据上面的分析,可以得到分类函数为:

$$P(C=c_1|X)=\frac{\sum_{s\in c_1}\exp(\lambda_s.F_s(S,X))}{\sum_{\forall c_i}\sum_{s'\in c_i}\exp(\lambda_{s'}.F_{s'}(S,X))} \quad (11)$$

其中 c_i 为第 i 个动作类别,X 为测试中的观察序列。当一个动作序列进入模型后,按照上面给出的参数训练与更新方法,可以得出最后的动作类别。

假设样本库中具有 C 个不同的类别,每个类别有$|S|$个状态,观察序列持续时间为 M,则式(11)的分母项的计算复杂度为 $O(C.M)$,这是在可计算范围内,且不失分类的准确性。

4 实验与分析

本文采用 CMU(http://mocap.cs.cmu.edu/search.html)视频数据库里的样本进行实验。这个数据库已经按人类动作类型进行了分类标注(每个视频分段是一个动作种类),可以用来进行分割和分类。本文的目的主要是验证算法在动作分类上的可行性及正确率,尤其是一些比较模糊的动作种类。选取 4 个动作种类(正常行走、慢步行走、漫步、跑步等)各 5 个视频段共 20 个视频段,其中每个种类选取 3 个做训练集,2 个做测试集。

在视频段中按 30 帧/s 的速度提取图片,延续 1 s,提取 30 张图片进行训练。图片提取出来后要进行预处理,利用统计背景提取及运动分割相结合的方法,进行前景与背景的分离。利用 50 维的形状上下文特征和不同尺度的边缘对特征,将图像存储成一个向量作为观察样本。

对每一个人类动作,建立一个 3 个状态的隐马尔可夫模

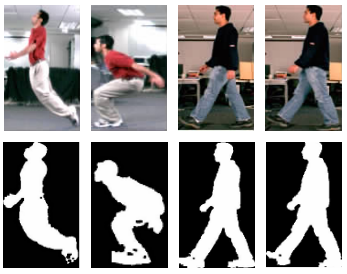

图 3　部分训练样本

(上面是样本,下面是对应的轮廓)

型,每一个状态使用一个高斯概率分布来表示。在训练过程中,利用 Viterbi 搜索算法分割出各状态所对应的结果,从而训练出每一个动作种类的每个状态。在分类测试过程中,利用 Viterbi 搜索算法将测试集分割出与模型最相近的路径,并计算相似度值,最后输出最高值者为分类结果。表 1 为不同算法分类正确率,从表中可以明显看出本文所提出的方法在分类正确率上优于 MEMM 方法,与带上下文窗口的 CRF 相当。

表 1　不同算法分类正确率比较

	分类正确率			
	正常行走	慢步行走	跑步	漫步
MEMM	16.3%	50.5%	75.0%	64.5%
CRF(w=0)	38.9%	86.5%	100.0%	65.0%
CRF(w=3)	100.0%	100.0%	100.0%	45.0%
文中所用方法	100.0%	95.0%	100.0%	76.5%

5　结论及未来工作

CRF 的架构可以随着应用层面的不同和变量之间实际相依性关系而有不同的图形结构。因为先后时间点的状态与观察样本相互影响而使得图形架构中含有回路,使得一般线性 CRF 推论方法在这里得不到有效运用,本文在充分理解联合树性质的基础上对具有回路的网状架构进行了改造,用构造联合树的方法来解决回路存在所产生的问题。利用观察样本与其相对应状态的相似度、状态转移概率作为特征函数来训练 HMCRF 模型参数。在对 CMU 的人类动作图像数据库的实验中,发现这样的架构在模型的训练收敛性与分类正确率上优于一般CRF 与 MEMM 方法。当然,还遗留一些问题有待进一步研究,例如:当考虑更长时间点相关时,接口会更大,计算更复杂,模型训练时间如何优化;人类动作分类特征提取方法,是否可结合一些领域知识,等等。

参考文献:

[1] Yang J,Xu Y,Chen C S.Human action learning via Hidden Markov Model[J].IEEE Transaction on Systems man and Cybernetics,1997,27(1):34-44.

[2] Zhang X,Naghdy F.Human motion recognition through fuzzy Hidden Markov Model[C]//Proceedings of International Conference on Computational Intelligence for Modeling,Control and Automation,2005.

[3] Lafferty J,McCallum A,Pereira F.Conditional random fields:probabilistic models for segmenting and labeling sequence data[C]//ICML,2001.

[4] McCallum A,Freitag D,Pereira F.Maximum entropy Markov models for information extraction and segmentation[C]//ICML,2000.

[5] Sminchisescu C,Kanaujia A,Li Z,et al.Conditional models for contextual human motion recognition[C]//Proceedings of International Conference on Computer Vision,2005.

[6] Kumar S,Herbert M.Discriminative random fields:a framework for contextual interaction in classification[C]//ICCV,2003.

[7] Torralba A,Murphy K,Freeman W.Contextual models for object detection using boosted random fields[C]//NIPS,2004.

[8] Sutton C,Rohanimanesh K,McCallum A.Dynamic conditional random fields:factorized probabilistic models for labeling and segmenting sequence data[C]//Proceedings of ICML'2004,2004.

[9] Tang J,Hong M,Li J,et al.Tree-structured conditinal random fields for semantic annotation[C]//ISWC06,2006.

[10] Haykin S,Principe J C,Sejnowski T J,et al.New directions in statistical signal processing from system to brains[M].[S.l.]:MIT Press,2005.

[11] 统计模式识别[M].王萍,杨培龙,罗颖昕,等译.2 版.北京:电子工业出版社,2004.

[12] McCallum A.Efficiently inducing features of conditional random fields[C]//UAI,2003.

(上接 179 页)

上的要求。今后在进一步提高算法的运算速度,同时在结合曲线道路模型,对弯道进行识别等方面还需要进一步的研究。

参考文献:

[1] Wang Yue,Shen Dinggang,Teoh Eam Khwang.Lane detection and tracking using BOSnake[J].Image Vision Compute,2004,22(4):269-280.

[2] Ma Lin,Zheng Nan-Ning.Scene slice analysis based lane detection and tracking[C]//2005 IEEE.

[3] 陆建业,杨明.Vision-based real-time road detection in urban traffic[C]//Proc SPIE Real-Time Imaging VI,Nasser Kehtarnavaz,2002,4666:75-821.

[4] Otsuka Y,Muramatsul S.Multitype lane markers recognition using local edge direction[C]//Proc of IEEE Intelligent Vehicle Symposium,2002.

[5] Turchetto R,Manduchi1 R.Visual curb localization for autonomous navigation[C]//IEEE/RST IROS'03,Las Vegas,October 2003.

[6] 程洪,郑南宁.基于主元神经网络和 K 均值的道路识别算法[J].西安交通大学学报,2003,37(8):812-8151.

[7] 刘加海,白洪欢,黄微凹.基于彩色和边缘信息融合的道路分割算法[J].浙江大学学报:工学版,2006,40(1).

[8] 王文明,赵荣椿.不同彩色空间的彩色图像边缘检测研究[J].计算机测量与控制,2006,14(12).

[9] 吕明忠,罗鹏,高敦岳.一种基于色差的彩色图像的边缘检测方法[J].华东理工大学学报,2001,27(5).

[10] 冈萨雷斯.数字图像处理[M].2 版.北京:电子工业出版社:470-489.

[11] Otsu N.A threshold selection method from gray-level histogram[J].IEEE Trans on SMC,1979,9:62-66.