

增量式人体姿态映射模型的学习方法

刘长红^{1,2} 杨扬¹ 陈勇³

(北京科技大学信息工程学院 北京 100083)¹ (江西师范大学计算机信息工程学院 南昌 330022)²
(南昌工程学院管理工程系 南昌 330099)³

摘要 判别式 3D 人体姿态估计方法直接学习图像观测到姿态之间的映射,需要大量训练集,而 GPR 对这种大训练集的映射模型学习由于计算复杂度太高而受到极大限制。提出了一种基于 GPR 和 LWPR 的增量式映射模型的学习方法,利用 GPR 学习各局部映射模型,基于 LWPR 的思想在线调整现有的模型和训练新的局部模型以及姿态估计。实验表明,该方法能够极大地减少大数据集上高斯过程回归的计算代价,并获得准确的姿态估计。

关键词 姿态估计,高斯过程回归,局部加权投影回归,增量学习

中图分类号 TP391.4 **文献标识码** A

Incrementally Learning Human Pose Mapping Model

LIU Chang-hong^{1,2} YANG Yang¹ CHEN Yong³

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(School of Computer Information and Engineering, Jiangxi Normal University, Nanchang 330022, China)²

(Department of Management Engineering, Nanchang Institute of Technology, Nanchang 330099, China)³

Abstract Discriminative approaches to 3D human pose estimation directly learn a mapping from image observations to pose, which requires large training sets. Gaussian process regression(GPR) to learn this mappings has been limited for high computational complexity, so we proposed a incrementally learning mappings based on GPR and Locally Weighted Projection Regression(LWPR). The approach utilized GPR to learn individual local models and LWPR to update existing models or learn a new local model for pose estimation. The experiment showed that the approach could greatly decrease computational complexity and exactly estimate the poses.

Keywords Pose estimation, Gaussian process regression, Locally weighted projection regression, Incremental learning

由于人体自身及其运动的高度复杂性、投影的二义性、遮挡等原因使得人体姿态估计非常具有挑战性。目前主要的方法有产生式方法和判别式方法。产生式方法是在姿态估计过程中使用先验知识(人体模型、运动模型),建立人体姿态的图像观测,然后通过相似性度量函数进行度量。为了找到与图像观测最匹配的姿态,需涉及高维姿态空间的搜索,而且需要进行初始化。而判别式方法直接学习图像观测到人体 3D 姿态的映射,不需初始化,计算复杂度低,一旦学习到这种映射,就能快速地估计姿态。Agarwal 和 Triggs^[1]使用侧影形状上下文特征,采用稀疏回归方法,如相关向量机(RVM)学习侧影形状到 3D 姿态之间的映射;Smichisescu 等^[2]基于条件贝叶斯混合专家模型,表示图像特征到 3D 姿态的多模映射,这些方法在相对少的动作集上能够进行准确的估计。然而要估计人体的各种复杂动作姿态,需要大量的训练样本,所学习的映射是多模、非线性映射。高斯过程回归(Gaussian Process Regression, GPR)^[3]为这种高维非线性映射函数的学习提供了一种强有力的方法,但 GPR 的计算复杂度非常高,为 $O(N^3)$ (N 为训练样本数),为了降低 GPR 的计算复杂度,提出

了各种方法,如稀疏高斯过程^[4,5]、混合专家模型^[6]。Bo 和 Smichisescu^[7]采用贪婪块坐标下降法(Greedy Block Coordinate Descent, GBCD)解决高斯过程回归在大数据集姿态估计中所带来的计算和存储负担;Urtasun^[8]采用混合局部高斯过程模型(Local Gaussian Process Models, LGP)的方法学习图像特征与姿态之间的映射,离线学习随机选择的 R 个邻域的 LGP 超参数,在线进行姿态估计,姿态估计过程采用多个专家分别预测,预测结果为多个专家预测值的加权之和。以上方法都是将各种运动作为一个整体来学习图像特征到 3D 人体姿态的映射模型,而人体运动的高度复杂性,运动姿态的多样性,使得很难一次性学习到完整的映射模型,所以本文提出一种基于 GPR 和局部加权投影回归(Locally Weighted Projection Regression, LWPR)^[9]的增量式学习方法,该方法能够不断地学习新运动的映射模型和估计姿态,而无须对已有模型再进行学习,因此极大地减少了计算代价。

1 GPR

给定训练数据集 $D = \{(x_i, y_i), i = 1, \dots, N\}$, x_i 为输入向

到稿日期:2009-04-08 返修日期:2009-07-01 本文受国家自然科学基金(60873192),江西省教育厅科技项目(GJJ09143)资助。

刘长红(1977—),女,博士生,讲师,主要研究方向为图像处理与模式识别,E-mail:ahong286@126.com;杨扬(1955—),男,博士,教授,主要研究方向为图像处理、模式识别、计算机网络、多媒体技术;陈勇(1973—),男,硕士,主要研究方向为电子商务、图像处理。

量, y_i 为目标值, 回归的目标就是从给定的样本中学习 x 和 y 之间的函数关系: $y_i = f(x_i) + \epsilon_i$, ϵ_i 是 0 均值和 σ_ϵ^2 方差的高斯噪声。高斯过程建模方法为回归问题提供了灵活的概率体系, GPR 就是假定函数 f 属于高斯过程:

$$p(f|X) = N(0, K) \quad (1)$$

其中, $f = [f_1, \dots, f_N]$ 是函数向量, $f_i = f(x_i)$, $X = [x_1, \dots, x_N]^T$ 和 K 是协方差矩阵, 其元素为协方差函数 $k(x_i, x_j)$, 协方差函数一般为核函数, 本文采用的协方差函数为:

$$k(x_i, x_j) = \beta_1 \exp\left(-\frac{\beta_2}{2} \|x_i - x_j\|^2\right) + \frac{\delta_{x_i, x_j}}{\beta_3} \quad (2)$$

高斯过程模型由它的协方差函数确定, 给定训练样本的高斯过程学习实际上是一个求解协方差函数中超参数的问题, 这些超参数可通过最小化负对数似然函数求解:

$$-p(X, \beta|Y) = \frac{D}{2} \ln|K| + \frac{1}{2} \text{tr}(K^{-1}YY^T) + C \quad (3)$$

其中, $Y = [y_1, \dots, y_N]^T$, C 是常量, D 是输出的维数, 那么上式方程可通过一般优化算法(拟牛顿法、共轭梯度法等)来求解, 结果即为协方差超参数。对于新数据点 x_* , GPR 给出了其目标值的均值 $f(x_*)$ 和方差 $V(x_*)$ 为:

$$f(x_*) = k^T(K + \sigma_n^2 I)^{-1}Y \quad (4)$$

$$V(x_*) = k(x_*, x_*) - k^T(K + \sigma_n^2 I)^{-1}k$$

其中, k 是训练样本集 X 与 x_* 的协方差向量, K 是训练样本集 X 的协方差矩阵, $k(x_*, x_*)$ 是 x_* 的自方差, Y 是训练样本目标值向量。

2 增量式学习框架

判别式人体姿态估计方法直接学习图像特征到 3D 姿态的高维非线性映射模型, 需要大量的训练样本集, 而非线性回归方法 GPR 的计算复杂度为 $O(N^3)$, 存储量为 $O(N^2)$, 因而在大训练集中的应用受到极大限制, 所以本文提出一种基于 GPR 和 LWPR 的增量式学习框架, 首先将人体的各种运动划分为更小的子集, 每个子集作为一个局部域, 利用 GPR 学习各局部映射模型, 然后基于 LWPR 的思想判断是在线调整现有的模型还是训练新的局部模型以及估计姿态。

LWPR 是一种增量式空间局部化的学习方法, 采用分段线性函数逼近, 将输入空间分为 K 个接受域, 对于每个训练数据点 (x, y) , 根据权值 w_k 进行局部化, 权值 w_k 通过高斯核计算:

$$w_k = \exp(-0.5(x - c_k)^T D(x - c_k)) \quad (5)$$

c_k 为每个局部域的中心, D 为核宽, 是一个对角矩阵; 查询点的预测为 K 个局部模型预测值的加权平均, 预测值 \hat{y} 为:

$$\hat{y} = \frac{\sum_{k=1}^K w_k * y_k}{\sum_{k=1}^K w_k} \quad (6)$$

其中, y_k 为第 k 个局部模型的预测值。

本文利用 LWPR 的增量式学习机制学习新运动的映射模型和估计姿态。与 LWPR 不同的是, 本文不是将各种运动数据集整体作为输入进行局部化和学习, 而是把每种运动作为一个局部域, 而且由于图像特征到 3D 姿态的映射是高维非线性的, 因此对于每个局部映射模型的学习采用 GPR 方法。整个过程包括 GPR 模型学习和姿态估计两部分, GPR 模型学习即学习各个局部 GP 模型超参数, 分为离线和在线, 对于一些已知运动, 分别离线学习其对应的局部 GP 模型并

存储, 在线过程包括新运动的 GP 模型学习和姿态估计。学习过程根据权值 w_k 来判定是否需要学习新的模型, w_k 采用协方差函数计算:

$$w_k = \exp\left(-\frac{\beta_2}{2} \|x - c_k\|^2\right) \quad (7)$$

如果 w_k 大于某个设定的值, 则选择最相近的模型, 修改该模型训练样本集的中心、协方差矩阵的逆和预测向量; 否则重新学习新的模型, 然后将其存入模型库。姿态估计则选择最相近的模型进行估计, 根据 w_k 的值选择最大的两个权值, 比较这两个权值, 如果相差很大, 则取最大权值对应的模型进行估计; 否则表示要估计的姿态是这两个模型之间的过渡姿态, 利用式(6)对这两个模型的估计值加权平均, 具体见算法 1 和算法 2。

算法 1 高斯过程回归模型学习

离线学习: 已知运动的局部 GP 模型学习;

在线学习:

输入新数据点 (x, y) ;

循环: 对所有的模型

$$\text{计算 } w_k = \exp\left(-\frac{\beta_2}{2} \|x - c_k\|^2\right);$$

循环结束;

求最大的 w_k ; $v = \max(w_k)$;

如果 v 大于某个设定的值, 那么

将该点插入最近的模型并修改该模型的中心值、协方差矩阵的逆和预测向量;

否则

重新学习新的 GP 模型;

算法 2 姿态估计

输入图像特征 x ;

循环: 对所有的模型

$$\text{计算 } w_k = \exp\left(-\frac{\beta_2}{2} \|x - c_k\|^2\right);$$

循环结束;

取最大的两个权值, 假设为 w_p 和 w_q ;

如果 $|w_p - w_q| < \delta$, 则第 p 和 q 个模型分别计算预测值为 y_p 和 y_q , 最

后估计值 \hat{y} 为:

$$\hat{y} = \frac{w_p * y_p + w_q * y_q}{w_p + w_q}$$

否则

估计值 \hat{y} 即为最大的 w_k 对应模型的估计值;

3 计算复杂度分析

与全局 GPR 方法相比, 本文方法不是一次性对所有运动训练样本集进行学习, 而是对各种运动分别学习其局部 GP 模型, 计算复杂度为各个局部模型学习时间加上模型判断时间: $O(RS^3) + O(RN)$, R 为局部模型数, S 为局部模型的大小, N 为样本总数。当新的运动出现时, 只对新的运动进行学习, 而不用涉及已经学习的模型。对于姿态估计, 本文根据权值来选择属于哪个模型, 然后用最相近的一个或两个模型计算估计值, 复杂度为最相近的一个或两个局部模型估计时间加上模型选择时间: $O(S^2) | O(2S^2) + O(R)$ 。与全局高斯过程相比, 其计算复杂度大大降低了, 如表 1 所列。

表 1 计算复杂度比较

模型学习	姿态估计
------	------

全局 GPR	$O(N^3)$	$O(N^2)$
本文的方法	$O(RS^3)+O(RN)$	$O(S^2) O(2S^2)+O(R)$

4 实验与分析

本文的实验数据来自 Mocap 运动捕捉数据库^[10], 包括各种人体运动数据。训练集为侧影特征和 3D 姿态数据对, 如图 1 所示。人体三维姿态表示为 54 维关节角向量, 不考虑平移量。人体侧影来自 poser 渲染的侧影, 不受人体大小的影响, 使用形状上下文文^[11]描述子编码, 每个侧影采样 400 个轮廓点, 每个轮廓点编码为 60 维的向量, 然后对所有训练样本的形状上下文采用 k-means 算法进行聚类, 形成 200 个码字本, 根据生成的码字本对每个侧影进行矢量量化, 量化后结果为 200 维的向量作为输入。



图 1 人体侧影和 3D 姿态

实验分为模型学习和姿态估计两部分。每种运动序列取奇数帧作为模型学习和偶数帧作为姿态估计。模型学习先选取了走、跑、足球三种运动, 各取奇数帧中的 50 帧、100 帧和 150 帧作为训练样本, 即总共各为 150 帧、300 帧和 450 帧, 分别使用全局 GPR 方法和本文方法进行学习。图 2 给出了全局 GPR 方法和本文方法的计算时间, 实验在 CPU 1.6G、内存 1G 的 PC 机上进行。实验结果表明, 全局 GPR 的计算时间随样本数的增多快速增长, 而本文方法的计算时间与样本数呈线性关系, 从而大大降低了计算复杂度, 可在线实时训练模型。

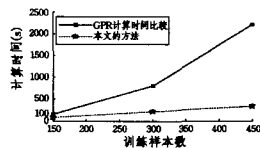


图 2 计算时间比较

然后上面 300 帧的基础上新增加 100 帧踢的动作训练, 全局 GPR 则要重新训练整个模型的 400 帧, 耗时 1029.71s, 而本文的方法仅仅只要训练增加的 100 帧, 耗时 53.42s, 远远低于全局 GPR 方法。

与模型学习相对应, 姿态估计则分别取走、跑、足球运动的偶数帧中的 50 帧、100 帧和 150 帧进行了实验, 分别求出关节角平均估计误差。表 2 给出了实验结果, 表明本文方法能够准确地估计出 3D 姿态, 甚至比全局 GPR 的性能更好一些。

表 2 3D 关节角平均估计误差(°)

	全局 GPR			本文的方法		
	走	跑	足球	走	跑	足球
50	1.75	0.92	2.0	1.38	0.9	1.67
100	1.78	0.96	2.47	1.27	0.92	2.25
150	1.96	0.87	2.41	1.28	0.87	2.27

结束语 本文基于 GPR 和 LWPR 的思想, 提出了一种增量式人体姿态映射模型学习方法, 该方法充分利用了非线性回归方法 GPR 在小样本集上能准确地学习映射模型和增量式学习方法 LWPR 的特性, 从而能够不断地对新运动人体姿态映射模型进行学习和估计。实验结果表明, 该方法极大地减少了大数据集上高斯过程回归的计算代价, 并能准确地估计出单帧图像的 3D 姿态。

参考文献

- [1] Agarwal A, Triggs B. Recovering 3D human pose from monocular images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2006, 28 (1): 44-58
- [2] Sminchisescu C, Kanaujia A, Li Z, et al. Discriminative density propagation for 3D human motion estimation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2005, 1: 390-397
- [3] Rasmussen C E, Williams C K. Gaussian Processes for Machine Learning[M]. Massachusetts Institute of Technology: MIT-Press, 2006
- [4] Candela J Q, Rasmussen C E. A unifying view of sparse approximate gaussian process regression[J]. Journal of Machine Learning Research, 2005: 1935-1959
- [5] Lawrence N D, Seeger M, Herbrich R. Fast sparse Gaussian process methods: The informative vector machine[M]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2003: 609-616
- [6] Rasmussen C E, Ghahramani Z. Infinite mixtures of gaussian process experts[J]. Advances in Neural Information Processing Systems, 2002, 14
- [7] Bo L-f, Sminchisescu C. Greedy block coordinate descent for large scale gaussian process Regression [C] // The Twenty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI'08. 2008
- [8] Raquel U, Trevor D. Sparse Probabilistic Regression for Activity-independent Human Pose Inference [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008
- [9] Vijayakumar S, D'Souza A, Schaal S. Incremental online learning in high dimensions[J]. Neural Computation, 2005, 17(12): 2602-2634
- [10] Mocap data[OL]. <http://www.mocapdata.com>
- [11] Belongie S, Malik J, Puzicha J. Shape Matching and Object Recognition Using Shape Contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2002, 24(4): 509-522

(上接第 244 页)

- [2] Cui Xining, Lü Quanyi. A parallel algorithm for block-tridiagonal linear systems[J]. Applied Mathematics and Computation, 2006, 173: 1107-1114
- [3] 谷同样, 刘兴平. 并行二级多分裂迭代方法[J]. 计算数学, 1998, 20: 153-166
- [4] 骆志刚, 李晓梅. 块三对角线性方程组的一种分布式并行算法[J]. 计算机学报, 2000, 23(10): 1028-1034
- [5] 盛跃宾, 宋晓秋, 刘德贵. 带状线性方程组的一种有效分布式并行算法[J]. 系统工程与电子技术, 2004, 26(7): 967-969

- [6] Varga R S. Matrix Iterative Analysis[M]. Englewood Cliffs, N. J: Prentice-Hall, 1962
- [7] 胡家麟. 线性代数方程组的迭代解法[M]. 北京: 科学出版社, 1999
- [8] Lv Quanyi, Xiao Manyu, Zhou Min. A parallel algorithm based on Galerkin theory for block-tridiagonal linear systems[J]. Applied Mathematics and Communication, 2007, 187(2): 1277-1285
- [9] Garey L E, Shaw R E. A parallel algorithm for solving Toeplitz linear systems[J]. Applied Mathematics and Communication, 1999, 100(2): 241-247