

基于带有噪声输入的稀疏高斯过程的人体姿态估计

夏嘉欣^{1,2} 陈曦³ 林金星⁴ 李伟鹏³ 吴奇^{1,2}

摘要 高斯过程回归 (Gaussian process regression, GPR) 是一种广泛应用的回归方法, 可以用于解决输入输出均为多元变量的人体姿态估计问题. 计算复杂度是高斯过程回归的一个重要考虑因素, 而常用的降低计算复杂度的方法为稀疏表示算法. 在稀疏算法中, 完全独立训练条件 (Fully independent training conditional, FITC) 法是一种较为先进的算法, 多用于解决输入变量彼此之间完全独立的回归问题. 另外, 输入变量的噪声问题是高斯过程回归的另一个需要考虑的重要因素. 对于测试的输入变量噪声, 可以通过矩匹配的方法进行解决, 而训练输入样本的噪声则可通过将其转换为输出噪声的方法进行解决, 从而得到更高的计算精度. 本文基于以上算法, 提出一种基于噪声输入的稀疏高斯算法, 同时将其应用于解决人体姿态估计问题. 本文实验中的数据来源于之前的众多研究人员, 其输入为从视频序列中截取的图像或通过特征提取得到的图像信息, 输出为三维的人体姿态. 与其他算法相比, 本文的算法在准确性, 运行时间与算法稳定性方面均达到了令人满意的效果.

关键词 姿态估计, 回归分析, 稀疏高斯过程, 噪声输入, 视频处理

引用格式 夏嘉欣, 陈曦, 林金星, 李伟鹏, 吴奇. 基于带有噪声输入的稀疏高斯过程的人体姿态估计. 自动化学报, 2019, 45(4): 693–705

DOI 10.16383/j.aas.2018.c170397

Sparse Gaussian Process With Input Noise for Human Pose Estimation

XIA Jia-Xin^{1,2} CHEN Xi³ LIN Jin-Xing⁴ LI Wei-Peng³ WU Qi^{1,2}

Abstract Gaussian process regression (GPR) is a common method for structured prediction and human pose estimation, in which input and output are both multivariate. Computational complexity is a significant consideration of GP regression and it can be reduced by sparse Gaussian algorithm. The fully independent training conditional (FITC) algorithm is a good method for sparse Gaussian process, and it can be applied to fully-independent input problems. Input noise is another significant consideration of GP regression. Moment matching can be used to solve trial input noise while training input noise can be modeled as output noise to achieve higher accuracy. On the basis of above algorithms, this study proposes a sparse Gaussian process with input noise for human pose estimation. A dataset from multiple people is used for experiments, in which the input is the image from video processing or image descriptor obtained by feature extraction, and the output is a three-dimensional human pose. The accuracy, runtime and stability of the algorithm are all satisfactory compared with other methods for human pose estimation.

Key words Human pose estimation, regression analysis, sparse Gaussian process (GP), noisy input, video processing

Citation Xia Jia-Xin, Chen Xi, Lin Jin-Xing, Li Wei-Peng, Wu Qi. Sparse Gaussian process with input noise for human pose estimation. *Acta Automatica Sinica*, 2019, 45(4): 693–705

收稿日期 2017-07-20 录用日期 2017-10-30
Manuscript received July 20, 2017; accepted October 30, 2017
国家自然科学基金 (61671293, 61473158, 51705242), 江苏省自然科学基金 (BK20141430), 上海浦江人才计划 (15PJ1404300), 浙江大学 CAD 和 CG 国家重点实验室开放课题 (A1713) 资助
Supported by National Natural Science Foundation of China (61671293, 61473158, 51705242), Natural Science Foundation of Jiangsu Province (BK20141430), Shanghai Pujiang Program (15 PJ1404300), and Open Project Program of the State Key Laboratory of CAD and CG, Zhejiang University (A1713)

本文责任编辑 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 上海交通大学电子信息与电气工程学院自动化系 上海 200240 2. 系统控制与信息处理教育部重点实验室 上海 200240 3. 上海交通大学航空航天学院 上海 200240 4. 南京邮电大学自动化学院 南京 210000

1. Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240 2. Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240 3. School of Aeronautics and Astronautics, Shang-

人体姿态估计是利用图像特征来估计各个个体部位在图像中的具体位置的研究过程^[1], 是一种应用广泛的结构化预测问题. 由于人体姿态估计在视觉跟踪、计算机动画^[2] 以及智能监控、虚拟现实^[3] 等领域均有广泛的应用, 因此, 在过去几十年中, 关于人体姿态估计问题的解决方法层出不穷.

高斯过程 (Gaussian process, GP)^[2–5] 及其变体^[6] 是一种典型且有效的人体姿态估计方法. 作为一种常见的基于贝叶斯概率论进行预测的回归方法, 高斯过程因其对非线性、高维、复杂和小样本问题的灵活性、有效性^[4] 及泛化能力强, 被广泛应用于各个领域. 然而, 人体姿态估计本身通常需要较大的

hai Jiao Tong University, Shanghai 200240 4. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210000

样本集, 所以简单地将高斯过程的基本模型应用于此类问题会耗费大量的运行时间与运算空间. 同时, 由于高斯模型本身对于噪声影响缺乏较高的鲁棒性, 而人体姿态估计等高维预测问题在实际应用中本身就将受到大量的噪声干扰, 传统高斯模型的预测准确性也将受到影响. 因此, 对于大样本问题预测的计算复杂度和对于噪声输入的预测准确性, 成为制约高斯过程应用于人体姿态估计问题的重要因素. 但是, 人体姿态估计作为一种非线性多元输入输出的拟合问题, 由于其输入输出变量的高维性, 该问题的预测模型本身就需要解决模型较为复杂、模型参数较多、超参数难以确定等种种问题. 而与当前更为主流的预测算法相比, 高斯过程具有模型容易实现, 参数复杂性低, 超参数自适应获取, 无需海量训练数据, 以及输出具有物理意义等突出优点, 这些优点注定了高斯过程在解决非线性高维预测问题方面的巨大优势与潜力. 因此, 若能同时改善计算复杂度与噪声鲁棒性这两项缺陷, 那么高斯过程将在解决人体姿势估计, 乃至所有多元输入输出问题方面, 发挥更加重要的作用, 产生更加深远的影响.

基于上述原因, 计算复杂度是高斯过程应用于人类姿态估计的一个重要的考虑因素. 稀疏高斯过程、混合专家模型^[7-8]、增量式学习^[9-10]方法等均是有效降低高斯过程计算复杂度的方法. 其中, 稀疏高斯过程最为常用. 该方法的本质为使用一组诱发输入 (Inducing input) 部分替代原输入进行假设^[11]与训练. 应用常规的高斯过程回归 (Gaussian process regression, GPR) 需要 $O(n^2)$ 的存储空间和 $O(n^3)$ 的运行时间, 其中 n 是训练点的数量. 由于人体姿态模型包含了数以千计的高维数据点, 因此其计算复杂度超乎想象. 然而, 如果运用稀疏高斯过程解决问题, 则存储空间将减少到 $O(nm)$, 而运行时间则将减少到 $O(nm^2)$, 其中 m 是诱发输入点的个数. 由于 $m \ll n$, 空间和时间成本得以显著降低. 本文将采用稀疏算法中的一种较为优秀的算法: 完全独立训练条件 (Fully independent training conditional, FITC)^[12]法, 来降低人类姿态估计问题中高斯回归的计算复杂度.

输入输出噪声是高斯算法的另一个重要考虑因素. 对于高斯回归的噪声研究始于很早之前, 但在大约十年前, 噪声才被扩展到应用于解决大规模和高维输入输出^[13]的问题. 对于测试输入样本的噪声, 可通过矩匹配的方法进行解决, 而对于训练输入样本的噪声, 先前的方法具有较高的计算复杂度. 为了解决这一问题, McHutchon 等^[14]在 2011 年提出了一种全新的算法, 本文将应用该方法来处理人体姿态估计问题中的输入噪声.

早在 1975 年, 人体关节模型就已经被证明包含

有丰富的信息, 并可用于行为识别. 此后的几十年, 众多研究人员均致力于三维人类姿态的重建^[15-16]与信息提取^[17-19], 几乎所有的经典模型如高斯过程、支持向量机、人工神经网络等, 都被成功且成熟地应用于人体姿态估计问题. 因此, 在近几年有关该问题的研究中, 更多的人着眼于通过提出更加新颖而复杂预测模型来提高人体姿态估计的预测精度^[3, 20-21], 或通过现有算法中加入新的优化模型形成混合模型来提升算法性能^[1, 22], 但着眼于从单个已成熟模型本身的数学机理出发, 通过改进其数学模型来提高人体姿态估计算法性能的研究则少之又少. 模型的数学原理是预测模型的根基, 从经典模型的数学原理角度对模型进行优化, 不但保证了模型的稳定性与可推广性, 还可以达到比参数优化更好更彻底的优化效果. 因此, 本文从高斯过程本身的数学原理出发, 利用稀疏算法与带有噪声输入的算法这两种优化方法进行模型改进, 并将改进后的模型应用于三维人体姿态估计问题, 在降低计算复杂性的同时, 也使得算法在输入噪声的影响下获得更好的预测精度. 相较于现有算法而言, 本文算法具有更强大的数学基础与泛化能力, 可以推广应用于多种结构化预测问题.

本文结构安排如下: 第 1 节对人体姿态估计问题进行描述; 第 2 节介绍了高斯过程的模型, 包括标准高斯过程; 第 3 节介绍带噪声输入的稀疏高斯算法, 同时对该算法的合理性和优越性进行简单验证. 第 4 节将本文算法与其他预测效果较好的人体姿态估计算法应用于 HoG 样本集^[17], 并对预测结果进行比较与评估. 第 5 节为结论.

1 问题描述

人体姿态估计是人体动作与行为的识别与分析的一个基础问题^[22]. 由于其在人机交互^[23]、人体活动分析和视频监控等领域都有着广泛的应用前景^[20], 因此关于人体姿态估计的方法也层出不穷. 在现有的人体姿态估计的算法中, 基于模型的方法更被研究者关注. 该方法的主要思路是通过建立数学模型来描述人体特征与空间位置, 并度量人体部位可能定位区域与真实人体部位外观的相似程度, 从而获得标准的输入输出集, 然后设计推理算法来确定相似度较高且符合人体模型约束的各部位定位区域^[1].

在实际问题中, 实验个体所处环境往往比较复杂, 会很大程度上增加外观模型的建立难度, 因此, 选取适当且有效的图像特征来简化模型建立过程十分重要. 本文考虑选取比较典型的方向梯度直方图特征 (Histogram of oriented gradients, HoG). HoG 特征计算的主要过程为: 1) 计算图像每个像素的梯

度大小和方向, 划分图像; 2) 将若干像素组成一个单元 (Cell), 统计每个单元的梯度直方图, 形成每个单元的特征描述; 3) 将若干单元组成一个块 (Block), 将每个块内所有单元的特征描述串联, 得到每个块的 HoG 特征描述; 4) 将图像内的所有块的 HoG 特征描述串联, 得到整个图像的 HoG 特征. HoG 不是考察图像的单个像素的特征, 而是通过计算局部区域中的定向梯度直方图以形成图像特征, 因此对光线和小幅度的位置偏移并不敏感.

由于 HoG 特征在人体姿态估计领域具有十分优秀的特征描述能力, 近年来使用 HOG 特征实现人体检测与估计系统成为研究热点^[24]. 本文的研究重点在于预测算法的提出与优化, 而非特征模型的建立, 因此, 本文借用 Poppe^[17] 基于 Sigal 等提出的 HumanEva-I^[16] 数据库建立的人体姿态 HoG 特征样本集进行仿真实验与分析.

HumanEva-I 是一个由视频序列组成的数据库, 该数据库中的视频序列由 3 个不同视角下的摄像机获取, 包含 4 名受试者的 6 种常见的行为.

在本文算法中, 定义输入变量 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$, 输出变量 $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)^T$, 输入输出均为高维变量, 其中, N 为样本个数, \mathbf{x} 为 D 维向量, \mathbf{y} 为 E 维向量. HoG 特征输入样本集 \mathbf{X} 的建立过程如下: 以单个摄像机获取的单帧图片为例, 1) 将图片进行背景删除、二值化、提取兴趣区域 (Region of interest, ROI) 等预处理操作; 2) 将 ROI 区域划分为 5 行 6 列的 30 个网格, 每个网格统计出 9 个 HoG 特征, 并将 30 个网格的 HoG 特征串联, 得到单一视角下的 270 维 HoG 特征; 3) 将 3 个视角下的特征串联, 形成 810 维 HoG 特征描述, 该特征即为输入集 \mathbf{X} 下的单个样本 \mathbf{x} . 输出变量样本集 \mathbf{Y} 下的单个样本 \mathbf{y} 是由 20 个三维身体关节位置坐标组合成的 60 维向量, 代表在 \mathbf{x} 描述下的人体姿态模型.

2 高斯过程模型

2.1 高斯过程标准模型

高斯过程是一种基于贝叶斯线性回归产生的一种回归模型. 与其他常用的预测算法相比, 高斯过程具有容易实现, 超参数自适应获取等优点.

回归过程, 简单来说就是根据训练集 \mathbf{X}_m 与 \mathbf{Y}_m 之间的映射关系, 预测新的测试点 \mathbf{X}_* . 最有可能的对应输出值 \mathbf{Y}_* 的过程. 高斯回归是在贝叶斯线性回归的基础上, 把自变量空间通过核函数映射到高维空间, 从而得到更好的预测结果的过程.

在高斯过程回归中, 时间域上所有随机变量均服从高斯联合分布, 其性质完全由均值函数和协方

差函数确定. 在实际观测中, 观测值可能会受到噪声污染, 假设噪声变量服从 $\varepsilon \sim N(0, \sigma_N^2)$, 输出样本集 \mathbf{Y} 的先验分布可以表示为

$$\mathbf{Y} \sim N(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}) + \sigma_N^2 I_N) \quad (1)$$

因此, 已知的训练输出值 \mathbf{Y}_m 与未知的测试输出值 \mathbf{Y}_* 的关系可以表示为

$$\begin{aligned} \begin{bmatrix} \mathbf{Y}_m \\ \mathbf{Y}_* \end{bmatrix} &\sim N \left(\begin{bmatrix} m(\mathbf{X}_m) \\ m(\mathbf{X}_*) \end{bmatrix}, \right. \\ &\left. \begin{bmatrix} k(\mathbf{X}_m, \mathbf{X}_m) + \sigma_N^2 I_N & k(\mathbf{X}_m, \mathbf{X}_*) \\ k(\mathbf{X}_*, \mathbf{X}_m) & k(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) = \\ &N \left(\begin{bmatrix} m(\mathbf{X}_m) \\ m(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} K_{mm} + \sigma_N^2 I_N & K_{m*} \\ K_{*m} & K_{**} \end{bmatrix} \right) \end{aligned} \quad (2)$$

其中, \mathbf{X}_m 表示训练输入值, \mathbf{X}_* 表示测试输入值, $m(\mathbf{X})$ 表示 \mathbf{X} 的均值函数, $k(\mathbf{X}, \mathbf{X}')$ 表示 \mathbf{X} 与 \mathbf{X}' 协方差函数, K_{**} , K_{*m} , K_{m*} , K_{mm} 分别是 $K(\mathbf{X}_*, \mathbf{X}_*)$, $K(\mathbf{X}_*, \mathbf{X}_m)$, $K(\mathbf{X}_m, \mathbf{X}_*)$, $K(\mathbf{X}_m, \mathbf{X}_m)$ 的简写形式.

由于变量 \mathbf{Y}_m 已知, 因此由贝叶斯公式可知, 后验传递函数 $p(\mathbf{Y}_* | \mathbf{Y}_m)$ 为

$$p(\mathbf{Y}_* | \mathbf{Y}_m) = \frac{p(\mathbf{Y}_m, \mathbf{Y}_*)}{p(\mathbf{Y}_m)} = N(\mu_*, \Sigma_*) \quad (3)$$

$$\mu_* = m(\mathbf{X}_*) + K_{*m}(K_{mm} + \sigma_N^2 I_N)^{-1}(\mathbf{Y}_m - m(\mathbf{X}_m)) \quad (4)$$

$$\Sigma_* = K_{**} - K_{*m}(K_{mm} + \sigma_N^2 I_N)^{-1}K_{m*} \quad (5)$$

其中, $p(\mathbf{Y}_m, \mathbf{Y}_*)$ 为先验传递函数, $p(\mathbf{Y}_* | \mathbf{Y}_m)$ 为高斯函数.

由式 (3)~(5) 可知, 由于时间域上所有随机变量均服从联合高斯分布, 因此该高斯过程可以被均值函数 $m(\mathbf{X})$ 和协方差函数 $k(\mathbf{X}, \mathbf{X}')$ 共同唯一决定. 同时, $k(\mathbf{X}, \mathbf{X}')$ 被称为从低维空间到高维空间映射的核函数. 在高斯过程中, 常用的核函数有常数核 (Constant function)、线性核 (Linear function)、径向基核 (Radial basis function) 等, 本文采用的协方差函数为平方指数协方差函数 (Squared exponential correlation function), 即

$$k(\mathbf{X}, \mathbf{X}') = \lambda_f^2 \exp \left(-\frac{1}{2}(\mathbf{X} - \mathbf{X}')^T \Lambda^{-1}(\mathbf{X} - \mathbf{X}') \right) \quad (6)$$

其中, Λ 为方差尺度, λ_f^2 为信号方差. $\theta = \{\Lambda, \lambda_f^2, \sigma_N^2\}$ 为超参数集合, 一般通过极大似然法求得, 具体过程为:

1) 建立训练样本条件概率的负对数似然函数及其关于 θ 的偏导数.

$$L(\theta) = \frac{1}{2} \mathbf{Y}_m^T K_{mm}^{-1} \mathbf{Y}_m + \frac{1}{2} \ln |K_{mm}| + \frac{N}{2} \ln 2\pi \quad (7)$$

$$\frac{\partial L(\theta)}{\partial \theta_i} = \frac{1}{2} \text{tr} \left((\alpha \alpha^T - K_{mm}^{-1}) \frac{\partial K_{mm}}{\partial \theta_i} \right) \quad (8)$$

其中, \mathbf{Y}_m 为训练样本集, K_{mm} 为训练样本协方差函数, N 为训练样本数, i 为超参数集的第 i 个元素.

2) 采用梯度下降法对偏导数进行优化, 通过规定的迭代次数, 得到超参数的最优解.

2.2 稀疏高斯过程

稀疏高斯过程是通过选取一定数目的诱发输入点来部分代替原输入点进行训练, 从而降低运算复杂度的方法.

在本文实验中, 由于输出变量 \mathbf{y}_i ($1 \leq i \leq N$) 彼此之间相互独立, 符合完全独立训练条件 (FITC), 因此, 在各种变量假设均与前文相同的基础上, 同时假设 \mathbf{Y} 变量彼此之间相互独立, 即

$$p(\mathbf{Y}_i, \mathbf{Y}_j | \mathbf{Y}_u) = p(\mathbf{Y}_i | \mathbf{Y}_u) p(\mathbf{Y}_j | \mathbf{Y}_u) \quad (9)$$

其中, $i \neq j$, 且 \mathbf{X}_u 与 \mathbf{Y}_u 是诱发输入与输出. 该条件即为完全独立条件. 在该条件下, 先验传递函数 $p(\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_u)$ 可表示为

$$\mathbf{N} \left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \\ m(\mathbf{X}_*) \\ m(\mathbf{X}_u) \end{bmatrix}, \begin{bmatrix} K_{11} & \cdots & K_{1N} & K_{1*} & K_{1u} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ K_{N1} & \cdots & K_{NN} & K_{N*} & K_{Nu} \\ K_{*1} & \cdots & K_{*N} & K_{**} & K_{*u} \\ K_{u1} & \cdots & K_{uN} & K_{u*} & K_{uu} \end{bmatrix} \right) \quad (10)$$

其中, 对于每个 $1 \leq i \leq N$ 而言, $K_{ui} = k(\mathbf{Y}_u, \mathbf{y}_i)$, $K_{iu} = k(\mathbf{y}_i, \mathbf{Y}_u)$, $K_{i*} = k(\mathbf{y}_i, \mathbf{Y}_*)$, $K_{*i} = k(\mathbf{Y}_*, \mathbf{y}_i)$.

在此处, 定义 Q_{ac} 为

$$Q_{ac} = K_{ab} K_{bb}^{-1} K_{bc} \quad (11)$$

其中,

$$\begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \\ \mathbf{x}_c \end{pmatrix} \sim \mathbf{N} \left(\begin{bmatrix} \mathbf{m}_a \\ \mathbf{m}_b \\ \mathbf{m}_c \end{bmatrix}, \begin{bmatrix} K_{aa} & K_{ab} & K_{ac} \\ K_{ba} & K_{bb} & K_{bc} \\ K_{ca} & K_{cb} & K_{cc} \end{bmatrix} \right) \quad (12)$$

如果高斯随机变量 \mathbf{x}_a 与 \mathbf{x}_c 关于给定的 \mathbf{x}_b 条件独立, 则

$$Q_{ac} = K_{ac} \quad (13)$$

证明. 首先, 考虑 \mathbf{x}_a 与 \mathbf{x}_c 在给定的 \mathbf{x}_b 下的条件分布函数. 以 \mathbf{x}_a 为例:

$$\begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \sim \mathbf{N} \left(\begin{bmatrix} \mathbf{m}_a \\ \mathbf{m}_b \end{bmatrix}, \begin{bmatrix} K_{aa} & K_{ab} \\ K_{ba} & K_{bb} \end{bmatrix} \right) \quad (14)$$

由于均值向量与协方差矩阵均已知, 因此可以得到 \mathbf{x}_a 与 \mathbf{x}_b 的联合概率分布. 同时, 通过积分求得边缘分布. 因此, 变量 \mathbf{x}_a 在给定的 \mathbf{x}_b 下的条件概率的协方差为

$$K_{a|b} = K_{aa} - K_{ab} K_{bb}^{-1} K_{ba} \quad (15)$$

同理, 变量 \mathbf{x}_c 在给定的 \mathbf{x}_b 下的条件概率的协方差为

$$K_{c|b} = K_{cc} - K_{cb} K_{bb}^{-1} K_{bc} \quad (16)$$

求 \mathbf{x}_a 与 \mathbf{x}_c 在 \mathbf{x}_b 条件下的联合概率分布, 可得其协方差矩阵为

$$\begin{pmatrix} K_{aa} - K_{ab} K_{bb}^{-1} K_{ba} & K_{ac} - K_{ab} K_{bb}^{-1} K_{bc} \\ K_{ca} - K_{cb} K_{bb}^{-1} K_{ba} & K_{cc} - K_{cb} K_{bb}^{-1} K_{bc} \end{pmatrix}$$

因为 \mathbf{x}_a 与 \mathbf{x}_c 关于给定的 \mathbf{x}_b 条件独立, 所以 $K_{ac} - K_{ab} K_{bb}^{-1} K_{bc} = 0$, 即, $K_{ac} = K_{ab} K_{bb}^{-1} K_{bc}$. \square

通过上述证明过程, 可以将式 (10) 改写为

$$\mathbf{N} \left(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_N) \\ m(\mathbf{X}_*) \\ m(\mathbf{X}_u) \end{bmatrix}, \begin{bmatrix} K_{11} & \cdots & Q_{1N} & Q_{1*} & K_{1u} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ Q_{N1} & \cdots & K_{NN} & Q_{N*} & K_{Nu} \\ Q_{*1} & \cdots & Q_{*N} & K_{**} & K_{*u} \\ K_{u1} & \cdots & K_{uN} & K_{u*} & K_{uu} \end{bmatrix} \right) \quad (17)$$

其中, $Q_{ab} = K_{au} K_{uu}^{-1} K_{ub}$. a 和 b 为两个任意的下标名称.

为了降低矩阵求逆的计算复杂度, 需要用 \tilde{K}_{mm} 代替 K_{mm} , 其中

$$\tilde{K}_{mm} = Q_{mm} + \text{diag}\{K_{mm} - Q_{mm}\} = Q_{mm} + \Lambda_{mm} \quad (18)$$

$$\Lambda_{mm} = \text{diag}\{K_{mm} - Q_{mm}\} \quad (19)$$

则 \mathbf{Y}_u 的后验传递函数可以表示为

$$\tilde{\mu}_u = m(\mathbf{X}_u) + Q_{um} \tilde{K}_{mm}^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m)) \quad (20)$$

$$\tilde{\Sigma}_u = K_{uu} - K_{um} \tilde{K}_{mm}^{-1} K_{mu} \quad (21)$$

为避免直接计算 \tilde{K}_{mm} 的逆矩阵, 式 (20) 和式 (21) 可以改写为

$$\tilde{\mu}_u = m(\mathbf{X}_u) + \tilde{\Sigma}_u \tilde{K}_{uu}^{-1} K_{um} \tilde{\Lambda}_{mm}^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m)) \quad (22)$$

$$\tilde{\Sigma}_u = K_{uu} (K_{uu} + K_{um} \Lambda_{mm}^{-1} K_{mu})^{-1} K_{uu} \quad (23)$$

所以, \mathbf{Y}_* 的后验传递函数为

$$\tilde{\mu}_* = m(\mathbf{X}_u) + K_{*u} K_{uu}^{-1} (\tilde{\mu}_u - m(\mathbf{X}_u)) \quad (24)$$

$$\tilde{\Sigma}_* = K_{**} - K_{*u} K_{uu}^{-1} (K_{uu} - \tilde{\Sigma}_u) K_{uu}^{-1} K_{u*} \quad (25)$$

从上述等式可以看出, 我们需要计算一个 $n \times n$ 的矩阵 Λ_{mm} 与几个 $m \times m$ 矩阵的逆矩阵. 由于 $m \ll n$, 并且 Λ_{mm} 为对角阵, 因此算法的计算复杂度得到了明显的改善.

如果假设输入变量 \mathbf{X} 服从零期望值的高斯分布, 则其后验分布函数可以表示为

$$\tilde{\mu}_* = m(\mathbf{X}_u) + K_{*u} K_{uu}^{-1} \tilde{\mu}_u \quad (26)$$

$$\tilde{\Sigma}_* = K_{**} - K_{*u} K_{uu}^{-1} (K_{uu} - \tilde{\Sigma}_u) K_{uu}^{-1} K_{u*} \quad (27)$$

2.3 带有噪声输入的高斯过程

在标准的高斯过程中, 常常存在两个假设: 其一是训练和用于测试的输入值 \mathbf{X}_m 和 \mathbf{X}_* 均为无噪声的; 其二是观测输出值 \mathbf{Y}_m 受到一个均值为零, 协方差为常数的高斯噪声 ε 的干扰. 但是在实际测量中, \mathbf{X}_m 和 \mathbf{X}_* 却并非是无噪声的. 也就是说, 输入值 X 并非某个确定的值, 而是一个随机变量.

2.3.1 测试输入点为随机变量

由前文可知, 当测试输入点 \mathbf{X}_* 为确定值时, 后验传递函数 $p(\mathbf{Y}_* | \mathbf{Y}_m)$ 表达式如式 (3) ~ (5) 所示.

但现在假设 \mathbf{X}_* 并非确定值, 而是一个服从高斯分布的随机变量, 即

$$\mathbf{X}_* \sim N(\hat{\mathbf{X}}_*, \Sigma_{X_*}) \quad (28)$$

若在此条件下得到 \mathbf{Y}_* 的后验分布, 最直观的方法是运用边缘概率密度与积分进行计算. 但是, 在这一过程中, 不但需要计算带有逆矩阵的指数的积分, 同时以上计算过程的结果还由一个非线性的随机变量 \mathbf{X}_* 决定, 因此计算过程太过复杂, 无法直接计算出积分结果.

为了解决积分计算过于复杂的问题, 本文采用矩匹配 (Moment matching) 的方法, 规定 $p(\mathbf{Y}_*)$ 服从高斯分布, 通过求取其均值与方差来确定预测点 \mathbf{Y}_* 的值.

$$\begin{aligned} \mu_* &= \mathbb{E}[\mathbf{Y}_*] = \int_{-\infty}^{\infty} \mathbf{Y}_* p(\mathbf{Y}_*) d\mathbf{Y}_* \times \\ &\int_{-\infty}^{\infty} \int_{\mathbf{X}} \mathbf{Y}_* p(\mathbf{Y}_* | \mathbf{X}_*) p(\mathbf{X}_*) d\mathbf{X}_* d\mathbf{Y}_* \times \end{aligned}$$

$$\int_{\mathbf{X}} \left(\int_{-\infty}^{\infty} \mathbf{Y}_* p(\mathbf{Y}_* | \mathbf{X}_*) d\mathbf{Y}_* \right) p(\mathbf{X}_*) d\mathbf{X}_* \quad (29)$$

由式 (4) 可知, $\int_{-\infty}^{\infty} \mathbf{Y}_* p(\mathbf{Y}_* | \mathbf{X}_*) d\mathbf{Y}_*$ 即是 \mathbf{X}_* 为确定值时 \mathbf{Y}_* 的期望值, 将式 (4) 代入式 (29) 可得

$$\begin{aligned} \mu_* &= \int_{\mathbf{X}} (m(\mathbf{X}_*) + K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1} \times \\ &(\mathbf{Y}_m - m(\mathbf{X}_m))) p(\mathbf{X}_*) d\mathbf{X}_* \quad (30) \end{aligned}$$

为简化上式, 定义

$$\bar{m}_* = \int_{\mathbf{X}} m(\mathbf{X}_*) p(\mathbf{X}_*) d\mathbf{X}_*$$

$$\bar{K}_{*m} = \int_{\mathbf{X}} K_{*m} p(\mathbf{X}_*) d\mathbf{X}_*$$

则式 (30) 可以改写为

$$\mu_* = \bar{m}_* + \bar{K}_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m)) \quad (31)$$

若假设 $m(\mathbf{X}) = 0$, 且协方差函数 K 为式 (6) 形式的平方指数协方差函数, 则 \bar{K}_{*m} 可以表示为

$$\begin{aligned} \bar{K}_{*m} &= \lambda_f^2 \sqrt{\frac{|\Lambda_{\mathbf{X}}|}{|\Lambda_{\mathbf{X}} + \Sigma_{\mathbf{X}_*}|}} \exp\left(-\frac{1}{2} (\hat{\mathbf{X}}_* - \mathbf{X}_m)^T \times \right. \\ &(\Lambda_{\mathbf{X}} + \Sigma_{\mathbf{X}_*})^{-1} (\hat{\mathbf{X}}_* - \mathbf{X}_m)) \quad (32) \end{aligned}$$

观察上式可以发现, 若 \mathbf{X}_* 为确定值, 即 $\mathbf{X}_* = \hat{\mathbf{X}}_*$ 且 $\Sigma_{X_*} = 0$, 则 $\bar{K}_{*m} = K_{*m}$, 此时式 (31) 与式 (4) 相等.

2.3.2 训练输入点为随机变量

同第 2.3.1 节, 假设 \mathbf{X}_m 并非确定值, 而是一个服从高斯分布的随机变量, 即

$$\mathbf{X}_m \sim N(\hat{\mathbf{X}}_m, \Sigma_{X_m}) \quad (33)$$

若在此条件下运用边缘概率密度与积分计算 \mathbf{Y}_* 的后验分布, 则会遇到与第 2.3.1 节相同的问题, 计算过程太过复杂, 无法直接计算出积分结果. 假如将第 2.3.1 节中的矩匹配方法应用于此处, 可以得到

$$\begin{aligned} \mu_* &= \int_{\mathbf{X}} (m(\mathbf{X}_*) + K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1} \times \\ &(\mathbf{Y}_m - m(\mathbf{X}_m))) p(\mathbf{X}_m) d\mathbf{X}_m \quad (34) \end{aligned}$$

式 (34) 与式 (30) 的最大区别在于, 需要计算的是 $K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1}$ 关于 \mathbf{X}_m 的积分, 并非关于 \mathbf{X}_* 的积分. 求 $K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1}$ 关于 \mathbf{X}_* 的积分比较简单, 把 $K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1}$ 视为常数即可, 但求 $K_{*m} (K_{mm} + \sigma_N^2 I_N)^{-1}$ 关于 \mathbf{X}_m 的积分则需要求逆矩阵关于非线性变量的积分, 这一过程十分复杂, 因此第 2.3.1 节的算法在此并不适用.

为了解决训练集的噪声输入问题, 采用一种全新的方法进行计算, 该算法称为带有噪声输入的高斯过程 (Noisy input Gaussian process, NIGP), 主要思想是将训练集的输入噪声转化为输出噪声, 运用常规的高斯过程回归解决问题.

首先做如下假设: 对于输入样本集 \mathbf{X}_m 与输出样本集 \mathbf{Y}_m , 假设其中的单个样本 \mathbf{x} 与 \mathbf{y} 分别为真值 $\hat{\mathbf{x}}$ 与 $\hat{\mathbf{y}}$ 在噪声影响下的测量值, 由此可得

$$\mathbf{y} = \hat{\mathbf{y}} + \boldsymbol{\varepsilon}_y \quad (35)$$

$$\mathbf{x} = \hat{\mathbf{x}} + \boldsymbol{\varepsilon}_x \quad (36)$$

$$\mathbf{y} = f(\mathbf{x} - \boldsymbol{\varepsilon}_x) + \boldsymbol{\varepsilon}_y \quad (37)$$

其中, $\boldsymbol{\varepsilon}_y \sim N(0, \boldsymbol{\Sigma}_y)$, $\boldsymbol{\varepsilon}_x \sim N(0, \boldsymbol{\Sigma}_x)$. 不难看出, 式 (35) 与式 (1) 假设的条件相同, 式 (36) 扩大到整个样本集即为式 (33).

随后, 将 $f(\mathbf{x} - \boldsymbol{\varepsilon}_x)$ 在 \mathbf{x} 附近进行 Taylor 展开, 并只取到一阶导数, 即可得到关于噪声的一阶模型:

$$\mathbf{y} = f(\mathbf{x}) - \boldsymbol{\varepsilon}_x^T \boldsymbol{\partial}_f + \boldsymbol{\varepsilon}_y \quad (38)$$

其中, $\boldsymbol{\partial}_f$ 是关于一个输入值的 D 维向量, 因此定义 Δ_f 为关于 N 个输入值的 $N \times D$ 维矩阵. 由此得到 \mathbf{y} 的概率为

$$p(\mathbf{y}|f) = N(f, \boldsymbol{\Sigma}_y + \boldsymbol{\partial}_f \boldsymbol{\Sigma}_x \boldsymbol{\partial}_f^T) \quad (39)$$

将式 (39) 与先验传递函数进行联立, 并且扩展到整个样本集, 由此得到后验函数的均值和协方差为

$$\boldsymbol{\mu}_* = K_{*m}(K_{mm} + \boldsymbol{\Sigma}_y + \text{diag}\{\Delta_f \boldsymbol{\Sigma}_x \Delta_f^T\})^{-1} \mathbf{Y}_* \quad (40)$$

$$\boldsymbol{\Sigma}_* = K_{**} - K_{*m}(K_{mm} + \boldsymbol{\Sigma}_y + \text{diag}\{\Delta_f \boldsymbol{\Sigma}_x \Delta_f^T\})^{-1} K_{m*} \quad (41)$$

3 带有噪声输入的稀疏高斯过程

本文第 2.2 节介绍了稀疏高斯过程, 第 2.3 节介绍了带有噪声输入的高斯过程. 两种算法均具有不同的原理与推导过程, 本节要在验证其合理性的基础上, 将两种算法进行结合, 得到更为广泛的适用范围及更好的预测结果.

3.1 建议模型

由第 2.3 节的推导过程可知, 在稀疏高斯过程中, \mathbf{Y}_* 的后验分布的均值与协方差如式 (24) 和式 (25) 所示. 通过观察可以发现, 式 (24) 与式 (4) 的表达形式相同, 由此可以推断式 (24) 可直接运用式 (4) 的运算过程. 论证如下:

根据矩匹配算法原理, 将式 (24) 代入式 (30), 得

$$\boldsymbol{\mu}_* = \int_{\mathbf{X}} (m(\mathbf{X}_*) + K_{*u} K_{uu}^{-1} (\tilde{\boldsymbol{\mu}}_u - m(\mathbf{X}_u))) \times p(\mathbf{X}_*) d\mathbf{X}_* \quad (42)$$

由式 (22) 可知, $\tilde{\boldsymbol{\mu}}_u$ 取值与 \mathbf{X}_* 无关, 因此在 \mathbf{X}_* 求积分的过程中, $K_{uu}^{-1} (\tilde{\boldsymbol{\mu}}_u - m(\mathbf{X}_u))$ 可以与式 (30) 中的 $(K_{mm} + \sigma_N^2 I_N)^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m))$ 一样视为常数. 因此式 (24) 可直接运用式 (4) 的运算过程, 得到带有噪声输入的 \mathbf{Y}_* 的后验分布的均值.

$$\boldsymbol{\mu}_* = \bar{m}_* + \bar{K}_{*u} K_{uu}^{-1} (\tilde{\boldsymbol{\mu}}_u - m(\mathbf{X}_u)) \quad (43)$$

其中, $\tilde{\boldsymbol{\mu}}_u = \tilde{\boldsymbol{\Sigma}}_u \tilde{K}_{uu}^{-1} K_{um} \tilde{\Lambda}_{mm}^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m))$, $\bar{m}_* = \int_{\mathbf{X}} m(\mathbf{X}_*) p(\mathbf{X}_*) d\mathbf{X}_*$, 并且当假设 $m(\mathbf{X}) = 0$ 时, \bar{m}_* 同样取 0 值.

对于训练样本集 \mathbf{X}_m 在噪声影响下的结果应用于稀疏高斯过程, 原理与测试样本在噪声影响下的原理相同. 因此, 可由式 (40) 和式 (42), 得到带有噪声输入的稀疏高斯过程 (Sparse Gaussian process with input noise, SGPIN) 的预测输出值 \mathbf{Y}_* 的后验分布的均值.

$$\boldsymbol{\mu}_* = \bar{m}_* + \bar{K}_{*u} K_{uu}^{-1} (\tilde{\boldsymbol{\mu}}_u - m(\mathbf{X}_u)) \quad (44)$$

其中,

$$\tilde{\boldsymbol{\mu}}_u = m(\mathbf{X}_m) + \tilde{\boldsymbol{\Sigma}}_u \tilde{K}_{uu}^{-1} K_{um} (\tilde{\Lambda}_{mm} (\boldsymbol{\Sigma}_y + \text{diag}\{\Delta_u \boldsymbol{\Sigma}_x \Delta_u^T\})^{-1} (\mathbf{Y}_m - m(\mathbf{X}_m)))$$

如果假设输入变量 \mathbf{X} 服从零期望值的高斯分布, 则 $\boldsymbol{\mu}_*$ 可以改写为

$$\boldsymbol{\mu}_* = \bar{K}_{*u} K_{uu}^{-1} \tilde{\boldsymbol{\mu}}_u \quad (45)$$

其中,

$$\tilde{\boldsymbol{\mu}}_u = \tilde{\boldsymbol{\Sigma}}_u \tilde{K}_{uu}^{-1} K_{um} (\tilde{\Lambda}_{mm} (\boldsymbol{\Sigma}_y + \text{diag}\{\Delta_u \boldsymbol{\Sigma}_x \Delta_u^T\})^{-1} \mathbf{Y}_m)$$

而对于 \mathbf{Y}_* 的后验分布的方差, 由于其推导过程过于复杂且占用较大篇幅, 同时对预测结果没有实质性影响, 因此只列出推导结果:

$$\begin{aligned} \boldsymbol{\Sigma}_* &= \lambda_f^2 - \lambda_f^4 \sqrt{\frac{|\Lambda_{\mathbf{X}}|}{|\Lambda_{\mathbf{X}} + 2\boldsymbol{\Sigma}_{\mathbf{X}_*}|}} \times \\ &\sum_{i=1}^{n_u} \sum_{j=1}^{n_u} L_{ij} \exp\left(-\frac{1}{2}(\mathbf{x}_{u_i} - \mathbf{x}_{u_j})^T \times \right. \\ &(2\Lambda_{\mathbf{X}})^{-1}(\mathbf{x}_{u_i} - \mathbf{x}_{u_j})) \times \\ &\exp\left(-\frac{1}{2}\left(\frac{\mathbf{x}_{u_i} + \mathbf{x}_{u_j}}{2} - \hat{\mathbf{x}}_*\right)^T \times \right. \\ &\left.\left(\frac{1}{2}\Lambda_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{X}_*} + \boldsymbol{\Sigma}_y + \boldsymbol{\partial}_f \boldsymbol{\Sigma}_x \boldsymbol{\partial}_f^T\right)^{-1} \times \right. \end{aligned}$$

$$\left(\frac{\mathbf{x}_{u_i} + \mathbf{x}_{u_j}}{2} - \hat{\mathbf{x}}_* \right) - \mu_*^2 \quad (46)$$

其中,

$$L = K_{uu}^{-1}(K_{uu} - \Sigma_{uu})K_{uu}^{-1} - \alpha\alpha^T$$

$$\alpha = K_{uu}(\tilde{\mu}_u - m(\mathbf{X}_u))$$

3.2 样例测试

本文将稀疏高斯过程和有噪声输入的高斯模型相融合, 提出了一种全新的 SGPIN 算法. 在将该算法应用于人体姿态估计问题之前, 为了进一步验证该种融合的必要性与合理性, 采用一些简单的数据对 SGPIN 算法、稀疏高斯算法 (FITC)、有噪声输

入的高斯算法 (NIGP) 和常规高斯算法 (GP) 进行测试与评估.

如图 1 所示, 在定义域为 $[-5, 5]$ 、值域为 $[0, 2.5]$ 内随机生成一条输入输出曲线, 并将该曲线设定为待预测的理想目标曲线. 在曲线上均匀选取若干个点, 并人为添加高斯噪声, 组成带有噪声输入的训练集. 随后分别运用 GP, FITC, NIGP 和 SGPIN 四种算法进行预测, 得到预测曲线, 同时对每种算法的均方误差 (Mean squared error, MSE) 和预测时间进行比较与评估, 评估结果 (10 次重复实验取平均值) 如表 1 所示.

结合图 1 与表 1 中的信息可知, SGPIN 算法在预测准确度与运行时间方面均明显优于其他三种基础算法, 可见稀疏算法与去噪算法的结合确实可以

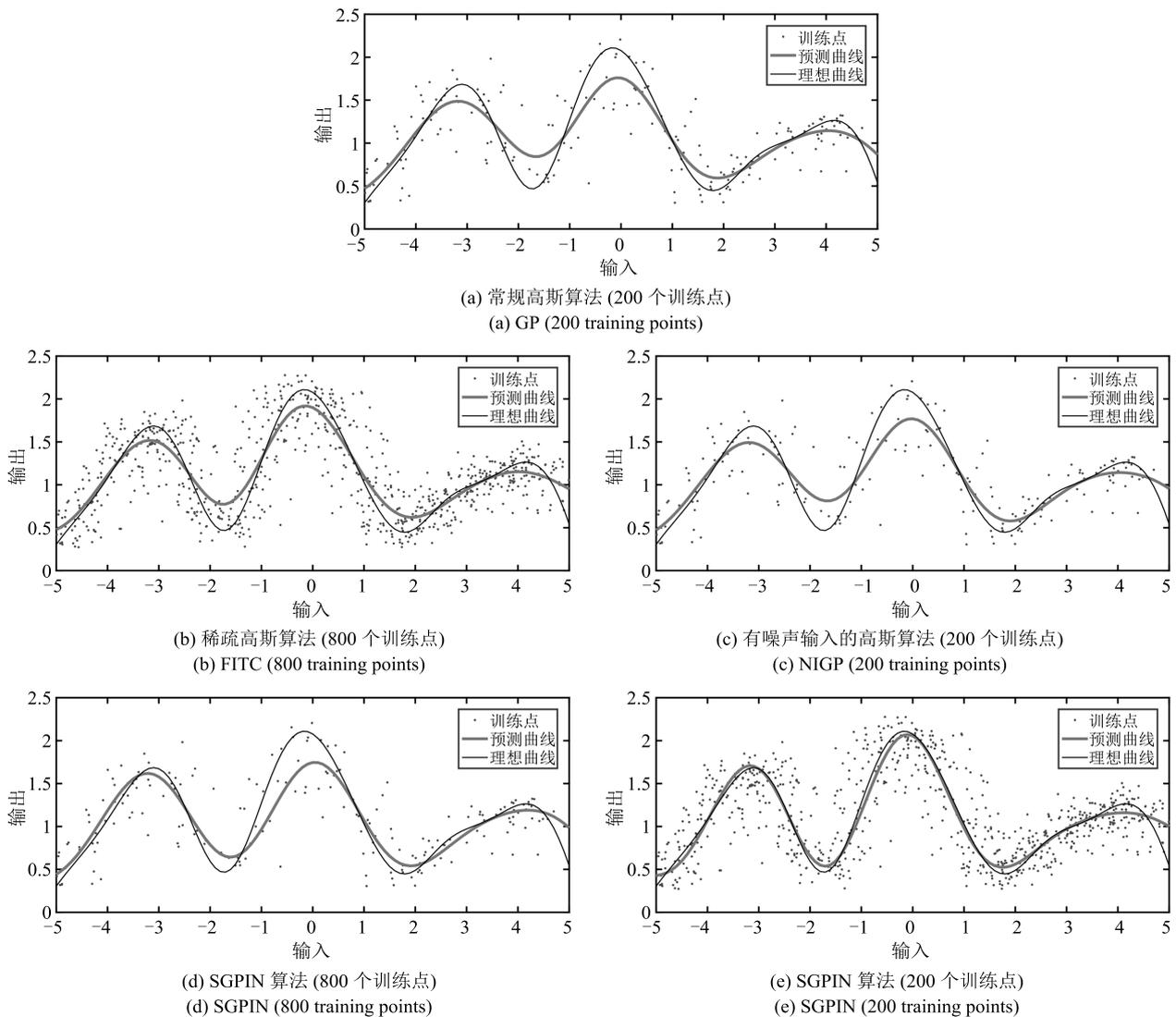


图 1 GP, FITC, NIGP 和 SGPIN 算法预测结果

Fig. 1 Predicting results of GP, FITC, NIGP and SGPIN

表 1 GP, FITC, NIGP 和 SGPIN 算法比较
Table 1 Comparison of GP, FITC, NIGP and SGPIN

算法	训练点个数	MSE (10^{-3})	运行时间 (s)
GP	200	31.1326	1.876034
FITC	800	18.6279	0.062001
NIGP	200	18.6279	13.630882
SGPIN	800	8.6265	0.003087
	200	18.4946	0.002612

大幅度提高高斯过程算法的性能, 达到更好更彻底的优化效果.

4 仿真研究与结果分析

本节主要对带有噪声输入的稀疏高斯 (SGPIN) 算法与双高斯过程算法 (Twin Gaussian processes, TGP)、 K -近邻算法 (K -nearest neighbor, KNN) 等预测算法进行比较与评估. 其中, TGP 算法是由 Bo 等^[25]提出的一种基于输入与输入两种联合概率分布的高斯过程, 在解决人体姿态估计问题时, 具有比 K -近邻算法、岭回归算法以及传统高斯过程等算法更高的预测准确度.

实验的数据集来自基于 HumanEva-I 数据库^[16]的人体姿态 HoG 特征集^[17], 包括 3 位测试对象的行走、慢跑、投掷捕捉, 做手势与拳击 5 个动作. 更多细节可以参考文献 [17].

4.1 输入、输出与误差度量

1) 表 2 为实验的数据集, 包含不用姿势和不同研究对象的样本个数 (每个视角下的单帧图片数). 对每个样本使用由 Poppe 计算的 810 维 HoG 特征描述^[17], 同时, 这个数据也用于评估双高斯过程和 KNN 等算法的性能^[25].

表 2 实验数据集
Table 2 Experimental set

特征	动作	个体 1	个体 2	个体 3	总数
HoG	Walking	1 176	876	895	2 947
	Jogging	439	795	831	2 065
	Throw/Catch	217	806	0	1 023
	Gestures	801	681	214	1 696
	Box	502	464	933	1 889
	Total	3 135	3 622	2 873	9 630

2) 输出变量是人体姿态, 表示为 20 个三维身体关节位置组合成的 60 维向量, 并且每个姿势都需要

进行预处理, 即将根关节位置设为原点, 其他关节的位置表示为与根关节的位置差. 这种标准化处理可以减少因研究对象的不同产生的不确定性.

3) 借用文献 [16] 提出的误差度量, 对于每一个输出向量, 误差公式可以表示为

$$D(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{M} \sum_{i=1}^M \| p_i(\mathbf{y}) - p_i(\hat{\mathbf{y}}) \| \quad (47)$$

其中, \mathbf{y} 是 60 维的估计输出, $\hat{\mathbf{y}}$ 是与之对应的真实输出, $M = 20$ 是每一个姿态中关节位置的个数, p_i 的输出值是第 i 个关节位置的三维向量. $\| \cdot \|$ 表示欧氏距离. 而对于整个输出序列, 平均误差可以表示为

$$Err = \frac{1}{N} \sum_{j=1}^N D(\mathbf{y}_j, \hat{\mathbf{y}}_j) \quad (48)$$

其中, N 表示输出向量的个数.

4.2 初始化与评估

1) 对于每一个数据集, 将其分成测试、诱发和训练三部分. 在 SGPIN 算法中, 通过迭代的梯度下降法确定超参数值, 这种方法会适当增加运行时间, 但是却可以得到更好的精度. 对于其他算法, 忽略诱发部分数据, 参考文献 [25] 进行参数值设定. 另外, 对于每种算法, 都会人为地在测试输入中添加高斯噪声, 同时添加一个极小的噪声矩阵来增加逆矩阵运算的稳定性.

2) 选取高斯过程 (GP)、双高斯过程 (TGP)、带有噪声输入的稀疏高斯过程 (SGPIN)、带有 KNN 算法的双高斯过程 (TGPKNN)、最优双核复合分类算法 (Kernel target alignment, KTA)^[26] 和希尔伯特施密特的独立性准则 (Hilbert-Schmidt independence criterion, HSIC)^[27] 下的 KNN 算法进行预测准确度和运行时间的评估. 在准确度方面, 比较每一个姿势的每一个输出向量的估计误差和不同姿势的平均估计误差. 同时, 为了保证评估的客观性, 每种算法都进行 5 次仿真运算, 取其平均值作为最终的运算结果.

表 3 列出了不同姿势的平均误差, 表 4 是每一种算法的运行时间. 这两个指标是评价算法综合能力的重要依据.

在表 3 和表 4 中, 所有模型均来自于 3 个研究对象的 5 个动作. 表中列出了样品数量, 最小误差和最短运行时间用粗体字显示, “/” 表示无数据集进行评估或样本集过小致使 KNN 算法无法在与其他样本相同 K 值下运行, 所有模型的诱发点个数均为 $n_u = 20$.

表 3 基于 HumanEva-I 数据集 HoG 特征的不同算法的平均误差

Table 3 Evaluation of average error of different algorithms based on HoG feature of HumanEva-I

研究个体	动作	样本数	GP	TGP	TGPKNN	KTA	HSICKNN	SGPIN
S1	Walking	1 176	398.5823	197.1179	193.9949	213.5265	218.6241	161.2112
	Jogging	439	383.7747	212.3234	212.2018	188.6683	196.0839	154.5919
	Throw/Catch	217	414.5873	174.2834	/	/	/	100.7592
	Gestures	801	415.3106	98.6237	102.5520	92.1541	156.6464	20.1770
	Box	502	426.6358	162.6801	163.3203	118.0500	149.5003	82.3949
S2	Walking	876	398.5817	197.1496	195.5694	206.7040	211.9735	160.4342
	Jogging	795	405.1201	213.0572	207.2430	227.3562	231.1777	176.1768
	Throw/Catch	806	421.5898	210.1543	199.3265	173.2717	189.7417	92.6742
	Gestures	681	410.0671	201.1053	201.7576	153.9103	173.0548	63.2473
	Box	464	421.3947	171.6007	109.1912	137.1031	159.5833	98.3920
S3	Walking	895	412.0019	219.2579	214.8589	236.1566	239.6487	177.3461
	Jogging	831	441.7053	211.1343	206.1400	233.5746	236.5287	184.2251
	Throw/Catch	0	/	/	/	/	/	/
	Gestures	214	473.7616	159.7482	/	/	/	40.3100
	Box	933	483.6534	214.1621	207.7578	186.5170	195.9815	120.6541
总数		9 630	284.0985	160.1196	162.0768	/	/	155.3066

表 4 基于 HumanEva-I 数据集 HoG 特征的不同算法的运行时间

Table 4 Evaluation of runtime of different algorithms based on HoG feature of HumanEva-I

研究个体	动作	样本数	GP	TGP	TGPKNN	KTA	HSICKNN	SGPIN
S1	Walking	1 176	0.11	26.77	24.67	28.16	27.87	18.02
	Jogging	439	0.03	8.47	10.43	10.18	10.26	21.65
	Throw/Catch	217	0.01	3.77	/	/	/	22.44
	Gestures	801	0.07	27.15	27.31	18.64	19.42	19.78
	Box	502	0.03	10.11	11.19	11.75	11.84	21.90
S2	Walking	876	0.08	20.86	25.83	20.03	20.26	22.04
	Jogging	795	0.07	18.06	17.86	17.64	17.74	23.32
	Throw/Catch	806	0.02	18.56	26.59	20.13	20.02	21.69
	Gestures	681	0.04	14.32	15.52	15.91	16.64	18.38
	Box	464	0.03	9.02	10.35	10.71	11.43	23.69
S3	Walking	895	0.09	22.78	22.63	20.75	20.95	21.13
	Jogging	831	0.08	21.83	20.13	18.51	19.01	20.36
	Throw/Catch	0	/	/	/	/	/	/
	Gestures	214	0.04	6.13	/	/	/	22.62
	Box	933	0.10	23.70	23.67	22.68	23.57	21.69
总数		9 630	11	1928	442	491	495	41

从表 3 可以看出所有算法的准确性. 其中, SGPIN 算法最为准确, 因为其具有比其他模型更小的平均误差. TGP 和 TGPKN 的误差在所有的运动中都是相似的, 因为这两种方法具有相同的核心算法. KTA 和 HSICKNN 算法与 TGP 算法的表现比较相似, 但 KTA 算法的预测准确性略优于 HSICKNN 算法. 常规 GP 模型与其他模型相比具有较大的误差, 因为它更受输入噪声的影响与干扰.

同时, 从表 3 可以看出, 在预测不同的姿态时, SGPIN 算法的优越性不尽相同. 在预测 Throw/Catch 和 Gestures 等变化比较强烈的姿态时, SGPIN 算法的预测准确度格外高, 在预测 Walking 和 Jogging 这类变化比较平缓的姿态时, 则表现了与其他算法相近的预测精度. SGPIN 算法的这种表现将在之后的算法稳定性分析中具体讨论.

表 4 的数据反映了所有算法的运行时间. 常规 GP 算法因为其模型的简单性, 比其他算法具有更少的运行时间. 另外 5 种算法在运行时间上没有较大的差别, 属于同一数量级. 观察表 4 中的每一列, 可以分析出样本数量对运行时间的影响. 具体表现为: 样品数量对 TGP, TGPKN, KTA 以及 HSICKNN 算法具有很大的影响, 对 GP 的影响较小. 在大多数情况下, 样本数越小, 平均误差也越小. 但是, SGPIN 算法几乎不受样本大小的影响, 因为在 SGPIN 算法中, 承担了主要运算量的诱发点数量不随样本数的变化而变化, 其取值均为 20.

接下来讨论 SGPIN 算法的稳定性. 图 2 和图 3 列举了在同一姿态中每一个输出向量的误差. 由于样本数量过大, 且预测误差在各样本子集的分布基本相同, 因此只截取每个样本集的 100 个样本进行描述. 图 2 给出了 SGPIN 算法与 TGP 和 TGPKN 算法的对比结果, 图 3 给出了 SGPIN 算法与 GP, KTA 和 HSICKNN 算法的对比结果.

从图 2 可以看出, 在很多样本中, TGP 和 TGPKN 的预测误差会随输出向量的不同产生很大的变化, 而 SGPIN 受到的影响较小. 此外, 若调细图 2 中的曲线, 可以发现 TGP 曲线存在大量毛刺, 而 TGPKN 和 SGPIN 的曲线则更为平滑. 原因是 TGP 算法利用了所有输入点的信息, 而 TGPKN 和 SGPIN 算法为了简化计算过程而忽略了一部分输入点.

从图 3 可以看出, GP 几乎不随输出向量的不同而产生变化, 原因是 GP 算法的原理比较简单, 注重的是整个样本集的总体信息. SGPIN 算法因为稀疏过程的存在, 输出向量对测量误差的影响也较小. 而 KTA 和 HSICKNN 算法与 SGPIN 算法相比则极不稳定, 预测准确度会随样本的不同产生巨大的波动, 而预测误差的离散程度也会随样本集的不同而

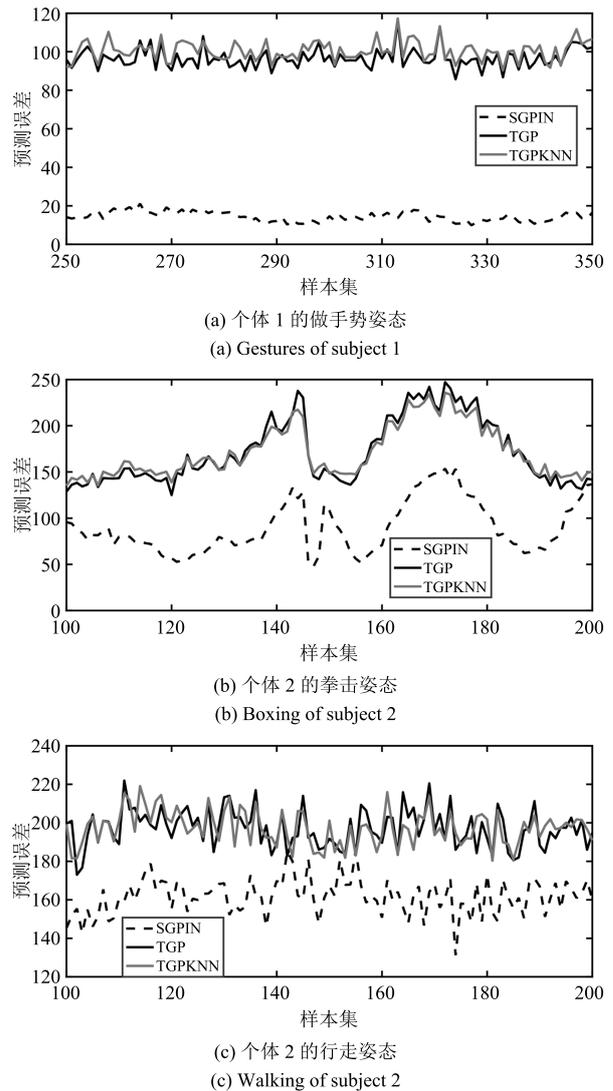


图 2 TGP, TGPKN 与 SGPIN 算法的误差比较

Fig. 2 Error comparison of TGP, TGPKN and SGPIN

有所不同.

结合图 2 和图 3 可以发现, 在预测 Walking 和 Jogging 这类变化比较平缓的姿态时, SGPIN 算法的稳定性与 TGP 和 TGPKN 算法基本相同, 且都弱于 GP 算法, 具体表现为预测曲线的波动程度相似. 在预测 Throw/Catch 和 Gestures 等变化比较强烈的姿态时, TGP, TGPKN, KTA 和 HSICKNN 算法的不稳定性使得算法产生了许多预测误差较大的样本, 极大影响预测精度, 这也是表 3 中 SGPIN 算法在预测变化比较强烈的姿态时准确度远远超过其他算法的原因.

另外, 在预测同一实验个体的同一种姿态时, 预测误差值的离散程度也反映了算法的稳定性. 由于在行走姿态的预测中 SGPIN 算法与其他算法的预测结果最为相似, 因此选取个体 3 的行走姿态, 重复进行 5 次仿真运算, 结果如表 5 所示.

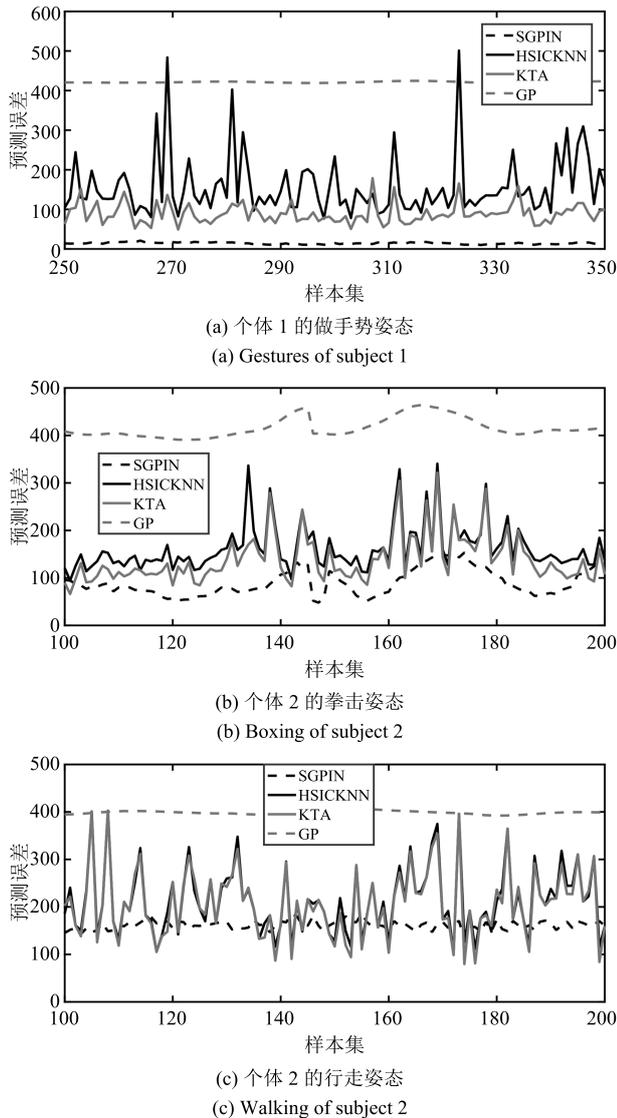


图3 GP, KTA, HSICKNN 与 SGPIN 算法的误差比较
Fig. 3 Error comparison of GP, KTA, HSICKNN and SGPIN

表5 个体3行走姿态的预测误差

Table 5 Predicting errors of subject 3 walking

	GP	TGP	TGPKNN	HSICKNN	KTA	SGPIN
1	412.0	219.1	214.8	236.3	239.6	177.6
2	412.0	218.0	214.9	232.5	235.7	176.9
3	412.0	220.2	214.6	237.1	240.9	177.7
4	412.0	220.4	215.6	241.6	244.8	177.5
5	412.0	218.7	214.5	233.4	237.3	177.0
方差	0.00	1.04	0.18	12.89	12.38	0.14

表5 最后一行的方差反映了预测误差值的离散程度. 方差越大, 算法的稳定性越低, 方差越小, 算法

的稳定性越高.

结合图2、图3和表5可以分析出6种算法的稳定性. HSICKNN和KTA算法极不稳定, TGP和TGPKNN算法稳定性一般, GP算法和SGPIN算法稳定性最好. 其中, SGPIN算法无论在预测变化较大的姿态方面, 还是在重复运行同一样本集时, 都能达到十分稳定的预测精度, 这两种性质也充分反映了SGPIN算法是一种能够推广到实际工程应用领域的人体姿态估计算法.

5 结束语

本文提出了一种从经典高斯过程模型的数学原理角度对模型进行优化来解决人体姿态估计问题的新思路. 算法的评估实验是基于HumanEva-I数据库的HoG特征集的三维人体姿态估计问题, 包括3个研究对象与5个人体姿态. 将本文算法与GP算法、TGP算法、TGPKNN算法、HSICKNN算法和KTA算法在预测准确度、运行时间和算法稳定性方面进行比较, 本文算法具有较为优秀的评估结果. 在算法准确度方面, SGPIN算法与其他算法相比, 具有较低的平均误差和较高的稳定性(较小的误差方差和较平滑的误差曲线), TGP和TGPKNN算法在稳定性方面表现一般, KTA算法和HSICKNN算法表现较差. 在运行时间方面, SGPIN算法并没有过于优异, 但是与其他算法相比其结果依旧是可以接受的. 同时考虑预测准确度、算法稳定性与运行时间三项因素, SGPIN算法是一种应用于人体姿态估计问题的更为有效的算法.

未来的工作首先是对于求解超参数的梯度下降算法的改进, 因为运用此种方法得到参数需要较长的运行时间, 且并未对输入信息进行充分利用. 此外, 需要更好地利用输入变量之间与输出变量之间的关系, 因为高维输入输出变量的各维度之间本身具有复杂的关联性. 同时, 将算法的应用领域进行拓展, 使其不仅局限于HoG这一特征, 甚至不仅局限于人体姿态估计这一问题, 而是用于解决更多的结构化预测问题.

References

- Shen Jian-Dong, Chen Heng. New human pose estimation algorithm based on HOG and color features. *Computer Engineering and Applications*, 2017, **53**(21): 190–194 (沈建冬, 陈恒. 融合HOG和颜色特征的人体姿态估计新算法. *计算机工程与应用*, 2017, **53**(21): 190–194)
- Wang J M, Fleet D J, Hertzmann A. Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(2): 283–298
- Yuan Zi-Hua, Li Feng, Zhou Shu-Ren. Human pose estimation based on Haar characteristics LBP texture feature.

- Computer Engineering*, 2015, **41**(4): 199–204
(袁紫华, 李峰, 周书仁. 基于 Haar 型 LBP 纹理特征的人体姿态估计. 计算机工程, 2015, **41**(4): 199–204)
- 4 Zhao X, Ning H Z, Liu Y C, Huang T. Discriminative estimation of 3D human pose using Gaussian processes. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, FL, USA: IEEE, 2008. 1–4
 - 5 Bratieres S, Quadrianto N, Ghahramani Z. GPstruct: Bayesian structured prediction using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(7): 1514–1520
 - 6 Ding M, Fan G L. Articulated Gaussian kernel correlation for human pose estimation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Boston, MA, USA: IEEE, 2015. 57–64
 - 7 Rasmussen C E, Ghahramani Z. Infinite mixtures of Gaussian process experts. In: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver, British Columbia, Canada: MIT Press, 2002. 881–888
 - 8 Yu Bin-Feng, Ji Hai-Bo. Sparse Bayesian mixture of experts and its application to spectral multivariate calibration. *Acta Automatica Sinica*, 2016, **42**(4): 566–579
(俞斌峰, 季海波. 稀疏贝叶斯混合专家模型及其在光谱数据标定中的应用. 自动化学报, 2016, **42**(4): 566–579)
 - 9 Liu Chang-Hong, Yang Yang, Chen Yong. Incrementally learning human pose mapping model. *Computer Science*, 2010, **37**(3): 268–270
(刘长红, 杨扬, 陈勇. 增量式人体姿态映射模型的学习方法. 计算机科学, 2010, **37**(3): 268–270)
 - 10 Yan Xiao-Xi, Han Chong-Zhao. Maximum likelihood estimation of multiple target states based on incremental finite mixture model. *Acta Automatica Sinica*, 2011, **37**(5): 577–584
(闫小喜, 韩崇昭. 基于增量式有限混合模型的多目标状态极大似然估计. 自动化学报, 2011, **37**(5): 577–584)
 - 11 Csató L, Opper M. Sparse on-line Gaussian processes. *Neural Computation*, 2002, **14**(3): 641–668
 - 12 Bijl H, van Wingerden J W, Schön T B, Verhaegen M. Online sparse Gaussian process regression using FITC and PITC approximations. *IFAC-PapersOnLine*, 2015, **48**(28): 703–708
 - 13 Snelson E, Ghahramani Z. Sparse Gaussian processes using pseudo-inputs. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: MIT Press, 2006. 1257–1264
 - 14 McHutchon A, Rasmussen C E. Gaussian process training with input noise. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain: ACM, 2011. 1341–1349
 - 15 HumanEva Dataset [Online], available: <http://humaneva.is.tue.mpg.de/>, November 3, 2017
 - 16 Sigal L, Balan A O, Black M J. HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, 2006, **87**(1–2): Article No. 4
 - 17 Poppe R. Evaluating example-based pose estimation: experiments on the HumanEva sets. In: Proceedings of the 2007 Computer Vision and Pattern Recognition Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2). Minneapolis, USA: IEEE, 2007.
 - 18 Su Ben-Yue, Jiang Jing, Tang Qing-Feng, Sheng Min. Human dynamic action recognition based on functional data analysis. *Acta Automatica Sinica*, 2017, **43**(5): 866–876
(苏本跃, 蒋京, 汤庆丰, 盛敏. 基于函数型数据分析方法的人体动态行为识别. 自动化学报, 2017, **43**(5): 866–876)
 - 19 Shakhnarovich G, Viola P, Darrell T. Fast pose estimation with parameter-sensitive hashing. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 750–757
 - 20 Han Gui-Jin, Zhu Hong. Human pose estimation algorithm based on pictorial structure model. *Computer Engineering and Applications*, 2013, **49**(14): 30–33
(韩贵金, 朱虹. 一种基于图结构模型的人体姿态估计算法. 计算机工程与应用, 2013, **49**(14): 30–33)
 - 21 Jiang H. Human pose estimation using consistent max covering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(9): 1911–1918
 - 22 Yang W L, Wang Y, Mori G. Recognizing human actions from still images with latent poses. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, USA: IEEE, 2010. 2030–2037
 - 23 Xu Feng, Zhang Jun-Ping. Facial microexpression recognition: a survey. *Acta Automatica Sinica*, 2017, **43**(3): 333–348
(徐峰, 张军平. 人脸微表情识别综述. 自动化学报, 2017, **43**(3): 333–348)
 - 24 Xu Yuan, Xu Xiao-Liang, Li Cai-Nian, Jiang Mei, Zhang Jian-Guo. Pedestrian detection combining with SVM classifier and HOG feature extraction. *Computer Engineering*, 2016, **42**(1): 56–60, 65
(徐渊, 许晓亮, 李才年, 姜梅, 张建国. 结合 SVM 分类器与 HOG 特征提取的行人检测. 计算机工程, 2016, **42**(1): 56–60, 65)
 - 25 Bo L F, Sminchisescu C. Twin gaussian processes for structured prediction. *International Journal of Computer Vision*, 2010, **87**(1–2): 28–52
 - 26 Cristianini N, Shawe-Taylor J, Elisseeff A, Kandola J. On kernel-target alignment. In: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver, British Columbia, Canada: MIT Press, 2001. 367–373
 - 27 Gretton A, Bousquet O, Smola A J, Schölkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. *Algorithmic Learning Theory*. Berlin Heidelberg, Germany: Springer-Verlag, 2005.



夏嘉欣 上海交通大学电子信息与电气工程学院自动化系硕士研究生. 2015 年获得上海交通大学学士学位. 主要研究方向为图像处理与机器学习.
E-mail: jessicax.1993@163.com
(**XIA Jia-Xin** Master student in the Department of Automation, School of Electronic Information and Electrical

Engineering, Shanghai Jiao Tong University. She received her bachelor degree from Shanghai Jiao Tong University in 2015. Her research interest covers image processing and machine learning.)



陈曦 上海交通大学航空航天学院讲师. 2014 年获得皇家墨尔本理工大学航空工程专业博士学位. 主要研究方向为故障预测与健康管管理, 机器学习, 结构健康监测. E-mail: chenxi1@comac.cc

(**CHEN Xi** Lecturer at the School of Aeronautics and Astronautics, Shanghai Jiao Tong University. He received

his Ph. D. degree in aerospace engineering from Royal Melbourne Institute of Technology University, Australia in 2014. His research interest covers prognosis and health management, machine learning, and structural health monitoring.)



林金星 南京邮电大学自动化学院副教授. 主要研究方向为复杂系统智能建模与控制, 切换奇异系统.

E-mail: jxlin2004@126.com

(**LIN Jin-Xing** Associate professor at the School of Automation, Nanjing University of Posts and Telecommunications. His research interest covers intelligent modeling and control of complex systems, switched

singular systems.)



李伟鹏 上海交通大学航空航天学院研究员. 2008 年获得哈尔滨工业大学硕士学位, 2011 年获得东京大学航空航天工程博士学位. 主要研究方向为湍流和气动噪声的数据挖掘.

E-mail: liweipeng@sjtu.edu.cn

(**LI Wei-Peng** Professor at the School of Aeronautics and Astronautics, Shanghai Jiao Tong University. He received his master degree from Harbin Institute of Technology in 2008 and his Ph. D. degree in aerospace engineering from University of Tokyo, Japan in 2011. His research interest covers data mining for turbulence, drag reduction, and noise control.)



吴奇 上海交通大学电子信息与电气工程学院自动化系副教授. 2009 年获得东南大学自动化学院控制工程与控制理论博士学位. 主要研究方向为深度多层网络建模与学习算法, 机器学习与模式识别. 本文通信作者.

E-mail: wuqi7812@sjtu.edu.cn

(**WU Qi** Associate professor in the Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University. He received his Ph. D. degree in control theory and control engineering from the School of Automation, Southeast University in 2006. His research interest covers deep multi-layers network modeling and learning algorithm, machine learning, and pattern recognition. Corresponding author of this paper.)