机器人运动轨迹的模仿学习综述

黄艳龙 [1] 徐 德 [2, 3] 谭 民 [3, 4]

摘 要 作为机器人技能学习中的一个重要分支, 模仿学习近年来在机器人系统中得到了广泛的应用. 模仿学习能够将人类的技能以一种相对直接的方式迁移到机器人系统中, 其思路是先从少量示教样本中提取相应的运动特征, 然后将该特征泛化到新的情形. 本文针对机器人运动轨迹的模仿学习进行综述. 首先详细解释模仿学习中的技能泛化、收敛性和外插等基本问题; 其次从原理上对动态运动基元、概率运动基元和核化运动基元等主要的模仿学习算法进行介绍; 然后深入地讨论模仿学习中姿态和刚度矩阵的学习问题、协同和不确定性预测的问题以及人机交互中的模仿学习等若干关键问题; 最后本文探讨了结合因果推理的模仿学习等几个未来的发展方向.

关键词 机器人技能学习, 模仿学习, 运动基元, 轨迹学习

引用格式 黄艳龙, 徐德, 谭民. 机器人运动轨迹的模仿学习综述. 自动化学报, 2022, **48**(2): 315−334

DOI 10.16383/j.aas.c210033

On Imitation Learning of Robot Movement Trajectories: A Survey

HUANG Yan-Long[1] XU De[2, 3] TAN Min[3, 4]

Abstract As a promising direction in the community of robot learning, imitation learning has achieved great success in a myriad of robotic systems. Imitation learning is capable of providing a straightforward way to transfer human skills to robots by extracting motion features from few demonstrations and subsequently employing them to new scenarios. This paper will review literature on trajectory learning by imitation for robots. The basic problems in imitation learning are first described in detail, such as skill adaptation, convergence and extrapolation. After that, state-of-the-art approaches are introduced, including dynamical movement primitives, probabilistic movement primitives and kernelized movement primitives. Later, various key problems are explained at length, e.g., learning of orientations and stiffness matrices, synergy and uncertainty prediction, as well as imitation learning in human-robot interaction. Finally, the possible future directions of imitation learning, for instance, the combination of imitation learning and causal inference, are discussed.

Key words Robot learning, imitation learning, movement primitive, trajectory learning

Citation Huang Yan-Long, Xu De, Tan Min. On imitation learning of robot movement trajectories: A survey. *Acta Automatica Sinica*, 2022, **48**(2): 315−334

机器人运动技能的模仿学习 (Imitation learning, IL), 又称示教学习 (Learning from demonstration, LfD) 或示教编程 (Programming by demonstration, PbD), 是指机器人通过学习示教样本来获得运动技能的一类算法, 其学习过程一般为从单个或少量示教轨迹中提取运动特征, 随后将该特征泛化到新的情形, 从而使得机器人具有较好的自适应性.

自 1999 年 Schaal[1] 提出机器人模仿学习的概念之后, 模仿学习作为机器人技能学习 (Robot learning) 领域中的一个重要分支近年来取得了许多重要的进展. 例如, Ijspeert 等[2] 提出了动态运动基元 (Dynamical movement primitives, DMP), 其仅需学习单条示教轨迹即可实现点到点和周期运动的泛化. 该方法利用弹簧阻尼模型和轨迹调整项, 可以在模仿示教技能时确保泛化轨迹收敛到目标点. Khansari-Zadeh 等[3] 提出了动态系统稳定估计 (Stable estimator of dynamical systems, SEDS), 该方法利用非线性求解器对多样本的高斯混合模型 (Gaussian mixture model, GMM) 的参数进行

收稿日期 2021-01-12 录用日期 2021-04-29
Manuscript received January 12, 2021; accepted April 29, 2021
国家自然科学基金 (61873266) 资助
Supported by National Natural Science Foundation of China (61873266)
本文责任编委 陈龙
Recommended by Associate Editor CHEN Long

1. 英国利兹大学计算机系 利兹 LS29JT 英国 2. 中国科学院自动化研究所精密感知与控制研究中心 北京 100190 中国 3. 中国科学院大学人工智能学院 北京 101408 中国 4. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 中国
1. School of Computing, University of Leeds, Leeds LS29JT, UK 2. Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China 3. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 101408, China 4. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

优化, 以使高斯混合回归 (Gaussian mixture regression, GMR)对应的自治系统 (即应用 GMR 预测状态变量对应的一阶微分, 如依据位置预测速度) 满足稳定性要求. Paraschos 等[4] 提出了基于高斯分布的概率运动基元 (Probabilistic movement primitives, ProMP), 其应用最大似然估计对轨迹参数的概率分布进行估计, 之后依据高斯条件概率的运算对轨迹进行泛化调整. Calinon 等[5] 提出了任务参数化高斯混合模型 (Task-parameterized GMM, TP-GMM), 该方法将训练轨迹投影到与任务相关的局部坐标系中并对变换后的相对运动轨迹进行概率建模, 克服了 GMM 在机器人任务空间中泛化的局限性. Huang 等[6] 提出了核化运动基元 (Kernelized movement primitives, KMP), 其通过对参数化轨迹和样本轨迹之间的 KL 散度 (Kullback-Leibler divergence) 进行最小化, 以及引入核技巧 (Kernel trick), 获得了非参的 (Non-parametric) 技能学习模型. 由于仅需要极少的样本即可实现对人类运动技能的迁移, 且无需其他先验知识或数据, 模仿学习被广泛应用于诸多领域, 如娱乐[7−10]、医疗[11−12]、护理[13−15] 和农业机器人[16]、仿人[17] 和外骨骼机器人[18−19] 以及人机交互[20−21] 等.

在上述运动轨迹的模仿学习之外, 模仿学习还包括其他的一些研究方向, 如行为复现 (Behaviour cloning, BC)[22]、直接策略学习 (Direct policy learning, DPL)[23] 和逆强化学习 (Inverse reinforcement learning, IRL)[24−25]. BC 和 DPL 在实质上可以理解为监督学习, 即学习示教样本中输入和输出的函数关系. 两者的区别是 DPL 在 BC 的基础上引入人类的交互反馈, 从而改进 BC 在长期规划中的不足, 特别是当训练和测试状态的概率分布存在显著差异的情形. IRL 假设训练样本中隐含的策略 (Policy) 在某种未知奖励函数 (Reward function) 下是最优的, 进而对奖励函数的参数进行优化, 最终在最佳奖励函数下应用强化学习 (Reinforcement learning, RL) 可求得该隐含的最优策略.

由于篇幅的限制, 本文仅针对机器人运动轨迹的模仿学习进行综述和讨论. 需要指出的是本文所讨论的模仿学习算法和 BC、DPL、IRL 存在着一定的差异. BC、DPL 和 IRL 主要侧重解决马尔科夫决策过程 (Markov decision process, MDP) 中的决策问题, 其中一个主要的特点是智能体 (Agent) 与环境存在交互且任意时刻的交互都会影响 MDP 下一时刻的状态, 这一过程常被描述为状态转换 (State transition). 轨迹的模仿学习侧重对运动轨迹的规划, 其输入通常为时间或其他无环境交互影响的状态[1]. 另外, 本文中涉及的一些算法如 GMR 和高斯过程 (Gaussian process, GP) 等可以划归到 BC 之中, 但考虑到这些方法的应用对象也包括机器人的轨迹学习, 因此我们仍将对其进行分析讨论.

之前的一些工作如文献 [26−27] 对模仿学习的部分问题进行了综述. 其中, 文献 [26] 仅介绍模仿学习中的少量工作, 未从算法的角度进行讨论. 文献 [27] 讨论了模仿学习中的任务参数化和轨迹协同两类问题, 但未涉及各种方法的具体推导思路. 不同于文献 [26−27], 本文主要综述机器人运动轨迹的模仿学习算法, 包括详细介绍模仿学习中的基本问题 (7 个) 和主要方法 (7 种), 以及着重讨论相关文献中的算法原理和该领域中存在的若干关键问题 (7 大类 11 小类).

本文的结构如下: 第 1 节对模仿学习中的一些基本问题进行描述, 随后在第 2 节中对几种主要的模仿学习算法进行介绍, 包括 GMM 和 GMR、GP、(半) 隐马尔科夫模型 (Hidden (Semi-)Markov Model, HMM/HSMM)、DMP、SEDS、ProMP 和 KMP. 第 3 节结合第 2 节的内容对模仿学习中的其他若干关键问题进行综述. 第 4 节对机器人轨迹模仿学习的未来发展方向进行探讨, 最后在第 5 节给出总结.

1 模仿学习中的一些基本问题

本节讨论模仿学习中的一些基本问题, 包括学习对象 (What to imitate)、技能复现 (Reproduction)、技能泛化 (Adaptation)、多轨迹的概率特征 (Probabilistic features)、收敛性 (Convergence)、外插 (Extrapolation)、时间输入和高维输入等问题.

1.1 模仿学习的对象

模仿学习具有广泛的适用范围, 如学习控制策略[1]、人对物体的操作策略[28] 以及人类的示教轨迹[3−4, 6] 或降维后的轨迹[29] 等. 考虑到本文的综述范围, 即适用于轨迹规划的模仿学习, 故仅以人类示教轨迹为学习对象进行分析.

目前常见的模仿学习是通过人类对机器人进行示教 (Kinesthetic teaching), 从而实现人类技能向机器人的迁移. 具体来说, 在重力补偿 (Gravity compensation) 模式下, 针对特定的任务人类可以直接地拖动机器人对其进行示教, 同时通过机器人自身的传感器、正向运动学以及视觉系统等记录机器人的关节角度、末端位置和姿态、力和力矩以及环境状态 (如物体或障碍物的位置、其他协作机器

[1] 在一些文献中轨迹的模仿学习被归类为 BC, 然而考虑到其研究内容的差异, 本文采用不同的划分方式.

人或用户的状态等), 进一步则利用模仿学习算法对经由示教所得的轨迹进行学习以达到对示教技能模仿的目的.

以图 1 为例, 在记录人类示教下机器人末端的位置和姿态后 (如第一行所示), 利用模仿学习算法可将学习到的技能应用的新的情形, 即生成新的末端位置和姿态轨迹 (如第二、三行所示). 图 2 给出了图 1 中粉刷任务对应的示教轨迹以及泛化轨迹, 其中圆圈表示泛化情形下的期望路径点. 需要说明的是, 除了对机器人进行拖动示教, 其他的方式还包括利用视觉捕捉系统[31−32] 采集人类的示教轨迹等.

图 1　KMP 在粉刷任务中的应用[30]. 第一行表示技能的示教, 第二行和第三行分别对应新情形下的泛化

Fig. 1　The application of KMP in painting tasks[30]. The first row illustrates kinesthetic teaching of a painting task while the second and third rows correspond to skill adaptations in unseen situations

1.2　技能的复现和目标点泛化问题

针对示教轨迹的学习, 首先需要考虑的是技能的复现和泛化问题. 前者是指模仿学习算法能够对示教轨迹进行准确地复现, 而后者则指学习算法将示教的技能应用到新的不同于示教的情形. 以图 3 为例, (a) 表示 DMP 的技能复现, 其中 DMP 生成的轨迹 (实线) 能够很好地重复示教轨迹 (虚线); (b)~(c) 均对应 DMP 的技能泛化, 其中 DMP 生成一条从新的起点 (圆圈) 收敛到新的目标点的轨迹, 该轨迹不同于示教轨迹. 在实际系统中, 技能的泛化问题是十分重要的. 以抓取为例, 技能泛化使得机器人在学习少量的示教轨迹之后, 能够对不同位置上的物体进行抓取而不需要新的示教样本.

1.3　带有中间路径点或速度要求的泛化问题

除了对于目标位置的泛化, 其他的泛化问题还包括经过期望的 (一个或多个) 中间路径点以及对位置和速度的同时泛化. 这里以打乒乓球机器人为例[33], 示教轨迹通常只包含少量的几条击打轨迹. 然而, 在实际系统中机器人应根据乒乓球的方位以及速度来调整其对应的击打位置和速度, 因此需要

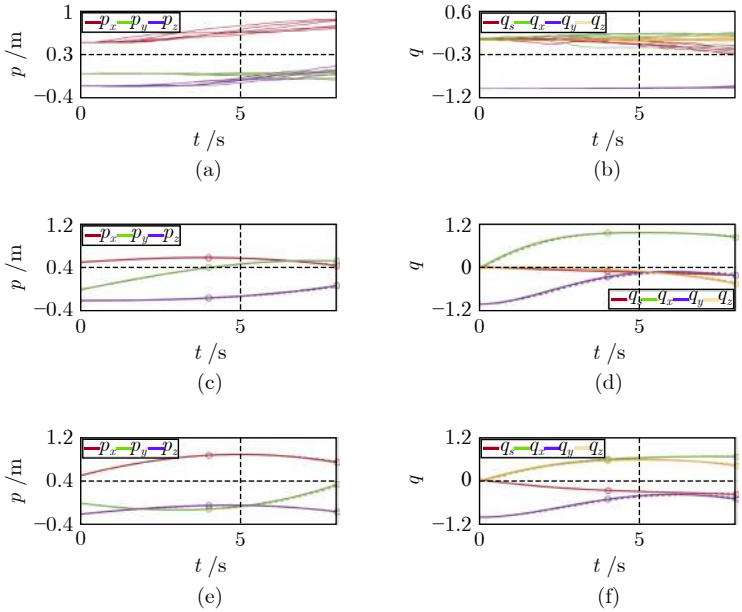

图 2　粉刷任务中的示教轨迹 (a)~(b) 以及泛化轨迹 (c)~(f), 其中 (c)~(d) 和 (e)~(f) 对应不同情形下的泛化[30]. $[p_x\ p_y\ p_z]^{\mathrm{T}}$ 和 $[q_s\ q_x\ q_y\ q_z]^{\mathrm{T}}$ 分别表示机器人末端的位置和四元数姿态. 圆圈为泛化时对应的期望路径点

Fig. 2　Demonstrations (a)~(b) and adapted trajectories (c)~(f) in painting tasks, where (c)~(d) and (e)~(f) correspond to different adaptations. $[p_x\ p_y\ p_z]^{\mathrm{T}}$ and $[q_s\ q_x\ q_y\ q_z]^{\mathrm{T}}$ denote Cartesian position and quaternion, respectively. Circles depict various desired points

考虑对位置和速度的同时学习和泛化. 图 4 为应用 KMP 对示教的书写技能进行泛化, 其中泛化后的轨迹能够经过新的起始、中间和目标点, 且每个期望点又包括期望的位置和速度.

1.4 多条示教轨迹的概率问题

在对人类的示教轨迹进行学习时, 需要考虑不同示教轨迹之间的差异. 以抓取为例, 即使针对同一个物体, 多次示教的轨迹仍可能存在不同程度的变化. 针对多条示教轨迹的问题, 需要考虑对轨迹中的概率分布进行学习. 这里仍以书写任务为例, 图 5 给出了应用 GMM 和 GMR 对多条轨迹进行概率学习的示意图, 其中 (d) 中的实线表示多条轨迹的均值而阴影部分的幅度则表征多条轨迹之间的变化程度.

1.5 收敛性问题

收敛性问题存在于基于动态系统 (Dynamical systems) 的模型中, 如学习轨迹中速度 $\dot{\boldsymbol{\xi}}$ 随位置 $\boldsymbol{\xi}$ 的变化趋势 (即学习 $\dot{\boldsymbol{\xi}}(t)=\boldsymbol{f}(\boldsymbol{\xi}(t))$ 对应的模型) 或学习轨迹中加速度 $\ddot{\boldsymbol{\xi}}$ 随位置 $\boldsymbol{\xi}$ 和速度 $\dot{\boldsymbol{\xi}}$ 的变化趋势 (即学习 $\ddot{\boldsymbol{\xi}}(t)=\boldsymbol{f}(\boldsymbol{\xi}(t),\dot{\boldsymbol{\xi}}(t))$ 对应的模型) 等. 以 $\dot{\boldsymbol{\xi}}(t)=\boldsymbol{f}(\boldsymbol{\xi}(t))$ 为例, 在利用模仿学习获得函数关系 $\boldsymbol{f}(\cdot)$ 之后, 可以根据当前的位置计算期望的速度从而能够计算出下一时刻期望的位置, 依此迭代下去即可获得完整的轨迹. 收敛性是指当 $t\to+\infty$ 时, $\boldsymbol{\xi}(t)$ 以零速度和零加速度收敛于期望的位置. 该特征可以有效地应用于当轨迹执行时存在较大干扰的情形. 收敛性也常常用于解决针对新目标点的泛化问题, 如图 3 中 DMP 即采用了稳定的二阶动态模型.

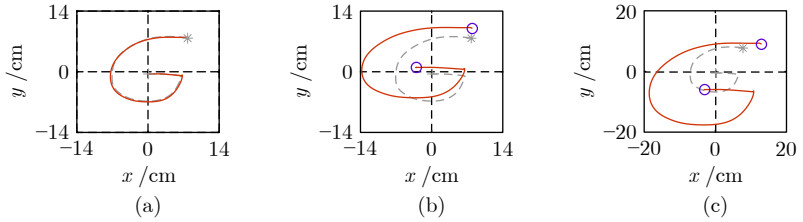

图 3 DMP 在书写字母中的应用. (a) 表示技能的复现, (b) ~ (c) 均表示技能的泛化, 其中实线对应 DMP 生成的轨迹, 虚线为示教轨迹并用 '*' 和 '+' 分别表示其起点和终点, 圆圈表示泛化轨迹需要经过的期望位置点

Fig. 3 The application of DMP in writing tasks. (a) corresponds to skill reproduction, (b) ~ (c) represent skill adaptations with different desired points. Solid curves are generated via DMP, while the dashed curves denote the demonstration with '*' and '+' respectively marking its starting and ending points. Circles depict desired points which the adapted trajectories should go through

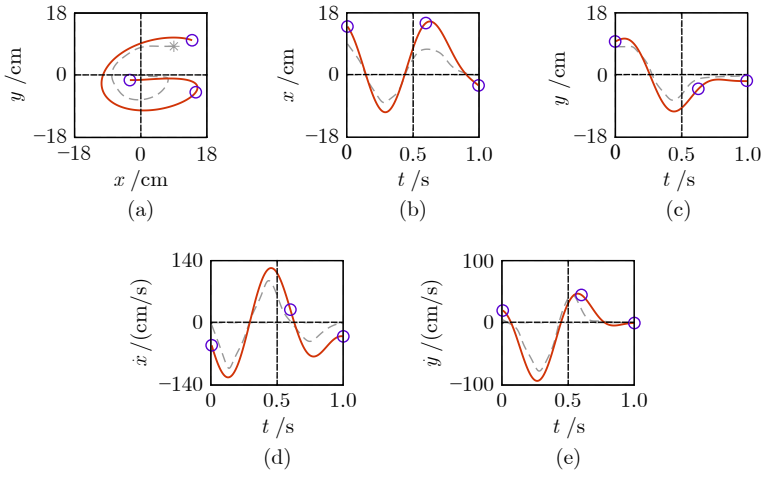

图 4 KMP 在书写字母中的应用. (a) 对应二维轨迹, (b)~(e) 分别表示轨迹的 x, y, $\dot{x}$ 和 $\dot{y}$ 分量. 实线对应 KMP 生成的轨迹, 虚线为通过 GMR 对示教轨迹进行建模得到的均值, 圆圈表示不同的期望点

Fig. 4 The application of KMP in a writing task. (a) plots the corresponding 2D trajectories, while (b)~(e) show the x, y, $\dot{x}$ and $\dot{y}$ components of trajectories, respectively. Solid curves are planned via KMP while the dashed curves are retrieved by GMR after modelling demonstrations. Circles denote various desired points

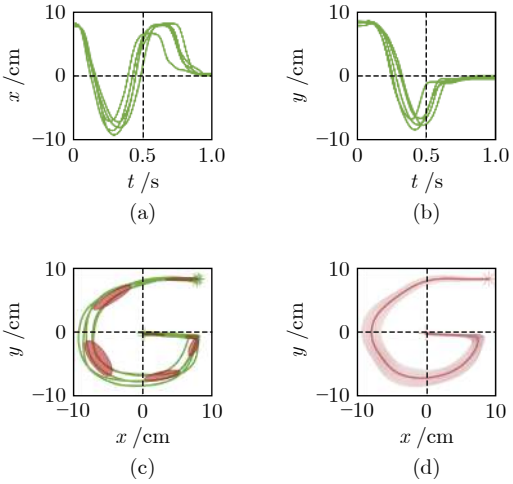

图 5　应用 GMM 和 GMR 对多条示教轨迹进行概率建模. (a) ~ (b) 分别对应示教轨迹的 x 和 y 分量, (c)~(d) 表示 GMM 和 GMR 的建模结果, 其中 (c) 中椭圆表示 GMM 中的高斯成分, (d) 中的实线和阴影部分分别表示多条轨迹的均值和方差

Fig. 5　The modeling of multiple demonstrations using GMM and GMR. (a)~(b) plot the x and y components of demonstrations. (c)~(d) depict the probabilistic features obtained via GMM and GMR, where the ellipses in (c) denote the Gaussian components in GMM, the solid curve and shaded area in (d) represent the mean and covariance of demonstrations, respectively

1.6　外插问题

外插问题是指将示教技能从整体上泛化到偏离示教区域的情形. 以物体搬运为例, 假设所有的示教轨迹都在用户的左侧, 具有外插特征的方法则允许将示教技能泛化到用户的右侧或其他远离示教区域的位置, 因此使得机器人具有更广泛的泛化能力. 图 6 为应用 DMP 进行外插的两个例子, 其中当期望的起始点和目标点整体远离示教区域时, DMP 依然能够生成与示教轨迹形状相似且经过期望点的轨迹.

1.7　示教轨迹的输入问题

在对示教轨迹进行学习时需要考虑对应的输入信息. 示教轨迹学习中的输入问题是指能否学习带有时间输入或高维输入的轨迹. 基于时间的技能学习能够在不同的时刻生成相应的轨迹, 如在某一个时刻到达某个期望的位置. 在上述的例子中, 图 2~6 均为学习时间驱动的轨迹. 针对高维输入的学习方法能够直接根据高维状态生成对应的轨迹. 如在图 7 所示的人机交互中, 当人的手部状态发生变化时, 机器人的轨迹也会立即作出相应的调整. 如人手的速度变快或变慢, 机器人的运动轨迹也会相应地变快或变慢. 该过程中的模仿学习可以理解为直接学习机器人和人之间的协调关系.

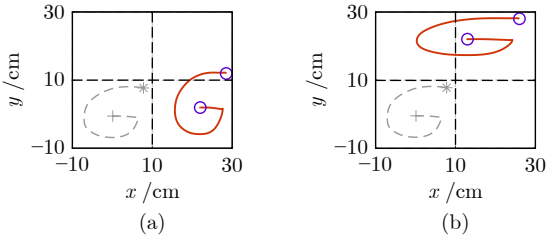

图 6　DMP 的外插应用

Fig. 6　The extrapolation application of DMP

图 7　KMP 在人机交互中的应用[34]. 第一行表示技能示教, 第二行为技能复现, 第三行对应新情形下的技能泛化

Fig. 7　The application of KMP in handover tasks[34]. The first row shows kinesthetic teaching of a handover task, while the second and third rows illustrate skill reproduction and adaptation, respectively

在本节最后, 需要指出的是当模仿学习应用于技能泛化或外插时, 仅需给定期望的任务目标即可生成相应的轨迹, 无需对轨迹形状进行几何分析或对轨迹进行分段处理等. 另外, 对于复杂的轨迹或存在高维输入时, 基于几何分析和变换的思路也是无法适用的. 在图 6 中, 给定期望的起始点和目标点, DMP 即能够直接生成保持示教轨迹形状的轨迹. 类似地, 在图 4 中给定期望时刻下对应的位置和速度 (图中对应 3 个期望点), KMP 即能直接生成满足要求的位置和速度轨迹, 而不需要其他任何中间步骤. 在图 7 中, 当人的手部状态 (高维变量) 发生变化时, KMP 即生成相应的机器人的轨迹.

2　几类主要的模仿学习方法

在获得示教轨迹之后, 需要对其进行相应的学习 (How to imitate). 机器人运动技能的模仿学习方法主要包括 GMM 和 GMR、HMM/HSMM、GP、DMP、SEDS、ProMP 和 KMP, 本节将结合第 1 节

中的基本问题对这些方法进行介绍和讨论.

2.1 高斯混合模型 (GMM) 和高斯混合回归 (GMR)

给定 M 条示教轨迹 $\{\{\boldsymbol{s}_{n,m},\boldsymbol{\xi}_{n,m}\}_{n=1}^{N_m}\}_{m=1}^{M}$, 其中 N_m 为第 m 条轨迹的长度, $\boldsymbol{s}\in\mathbf{R}^{\mathcal{I}}$ 表示 $\mathcal{I}$ 维输入信息 (如时间、位置或其他外部状态), $\boldsymbol{\xi}\in\mathbf{R}^{\mathcal{O}}$ 表示 $\mathcal{O}$ 维的轨迹变量, 如机器人末端位置、速度和加速度, 关节位置、速度和加速度, 以及力和力矩等. 两种典型的轨迹是: i) $\boldsymbol{s}$ 表示时间, $\boldsymbol{\xi}$ 为机器人末端位置、关节位置或力等,则示教轨迹表示时间驱动 (Time-driven) 的技能; ii) 如果 $\boldsymbol{s}$ 表示位置, $\boldsymbol{\xi}$ 为速度, 则示教轨迹对应自治的 (Autonomous) 动态系统.

GMM 可以对样本中输入和输出变量的联合概率分布 $\mathcal{P}(\boldsymbol{s},\boldsymbol{\xi})$ 进行建模, 即

$$\mathcal{P}(\boldsymbol{s},\boldsymbol{\xi})\sim\sum_{c=1}^{C}\pi_c\mathcal{N}(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c) \tag{1}$$

其中 C 为 GMM 中高斯成分的数量, π_c、$\boldsymbol{\mu}_c=\begin{bmatrix}\boldsymbol{\mu}_{s,c}\\ \boldsymbol{\mu}_{\xi,c}\end{bmatrix}$ 和 $\boldsymbol{\Sigma}_c=\begin{bmatrix}\boldsymbol{\Sigma}_{ss,c} & \boldsymbol{\Sigma}_{s\xi,c}\\ \boldsymbol{\Sigma}_{\xi s,c} & \boldsymbol{\Sigma}_{\xi\xi,c}\end{bmatrix}$ 分别表示第 c 个高斯成分的先验概率、均值和协方差. GMM 的参数可以通过期望最大化 (Expectation-maximization, EM) 算法 (文献 [35], 第 9.2 节) 进行迭代优化, 但需要事先指定高斯成分的数量. 常见的用于改进 GMM 参数估计的方法包括: i) 用 k 均值 (k-means) 对样本聚类, 然后用聚类结果初始化 GMM 的参数; ii) 结合贝叶斯信息判据 (Bayesian information criterion, BIC) 寻找最优的高斯数量[29]; iii) 贝叶斯 GMM (Bayesian GMM) 自动优化高斯成分的数量 (文献 [35], 第 10.2 节).

在得到 GMM 参数之后, 对于任意新的输入 $\boldsymbol{s}^*$ 均可利用 GMR 预测其对应轨迹 $\boldsymbol{\xi}^*$ 的条件概率分布, 即[34, 36−37]:

$$\mathcal{P}(\boldsymbol{\xi}^*|\boldsymbol{s}^*)=\sum_{c=1}^{C}h_c(\boldsymbol{s}^*)\mathcal{N}\left(\bar{\boldsymbol{\mu}}_c(\boldsymbol{s}^*),\bar{\boldsymbol{\Sigma}}_c\right) \tag{2}$$

其中

$$h_c(\boldsymbol{s}^*)=\frac{\pi_c\mathcal{N}(\boldsymbol{s}^*|\boldsymbol{\mu}_{s,c},\boldsymbol{\Sigma}_{ss,c})}{\sum\limits_{i=1}^{C}\pi_i\mathcal{N}(\boldsymbol{s}^*|\boldsymbol{\mu}_{s,i},\boldsymbol{\Sigma}_{ss,i})} \tag{3}$$

$$\bar{\boldsymbol{\mu}}_c(\boldsymbol{s}^*)=\boldsymbol{\mu}_{\xi,c}+\boldsymbol{\Sigma}_{\xi s,c}\boldsymbol{\Sigma}_{ss,c}^{-1}(\boldsymbol{s}^*-\boldsymbol{\mu}_{s,c}) \tag{4}$$

$$\bar{\boldsymbol{\Sigma}}_c=\boldsymbol{\Sigma}_{\xi\xi,c}-\boldsymbol{\Sigma}_{\xi s,c}\boldsymbol{\Sigma}_{ss,c}^{-1}\boldsymbol{\Sigma}_{s\xi,c} \tag{5}$$

进一步可以将式 (2) 近似为[34, 37]:

$$\mathcal{P}(\boldsymbol{\xi}^*|\boldsymbol{s}^*)\approx\mathcal{N}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) \tag{6}$$

其中

$$\hat{\boldsymbol{\mu}}=\sum_{c=1}^{C}h_c(\boldsymbol{s}^*)\bar{\boldsymbol{\mu}}_c(\boldsymbol{s}^*) \tag{7}$$

$$\hat{\boldsymbol{\Sigma}}=\sum_{c=1}^{C}h_c(\boldsymbol{s}^*)\left(\bar{\boldsymbol{\mu}}_c(\boldsymbol{s}^*)\bar{\boldsymbol{\mu}}_c^{\mathrm{T}}(\boldsymbol{s}^*)+\bar{\boldsymbol{\Sigma}}_c\right)-\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathrm{T}} \tag{8}$$

GMM 能够有效地学习多训练样本的概率特征, 包括时间输入和多维输入的情形. 然而, GMM 难以将其学习到的技能应用到与示教环境不同的情况. 为了改进 GMM 的自适应性 (即泛化能力), 常见的方法是应用强化学习, 如文献 [38] 利用行为评判算法[39] (Natural actor critic, NAC) 对 GMM 中高斯成分的均值进行优化. 由于需要大量的迭代优化, 这类方法不适用于在线技能的学习和调整.

2.2 (半) 隐马尔可夫模型 (HMM/HSMM)

HMM[40] 假设任意长度为 N 的观测序列 $\{\boldsymbol{o}_1,\boldsymbol{o}_2,\cdots,\boldsymbol{o}_N\}$ 是由 H 个隐含的未知状态 $\{s_1,s_2,\cdots,s_H\}$ 所产生, 同时假设当前时刻的观测值仅由当前时刻的隐含状态决定, 以及任意时刻的状态仅由其上一时刻的状态决定. 具体来说, HMM 包括三个主要要素 $\boldsymbol{\theta}=\{\pi_i,\{a_{i,j}\}_{j=1}^{H},b_i(\boldsymbol{o})\}_{i=1}^{H}$, 其中 π_i 为隐状态 s_i 的初始概率, $a_{i,j}$ 为隐状态从 s_i 转换到 s_j 的概率, $b_i(\boldsymbol{o})$ 表示当状态为 s_i 时观测到 $\boldsymbol{o}$ 的概率. 然而, 当对某个或某些隐含状态进行连续多次观察时, HMM 中状态频次的概率表征是不恰当的, 该概率会随连续观测频次的增加呈指数级下降. 例如, 对第 h 个隐状态连续观测 n 次 (即状态时长) 的概率为 $a_{h,h}^{n-1}(1-a_{h,h})$. 为了解决这一问题, HSMM[40] 对状态观测时长进行建模来取代 HMM 中的状态自循环, 其参数主要包括 $\{\{\pi_i,a_{i,j},b_i(\boldsymbol{o}),c(s_i)\}_{i=1}^{H}\}_{j=1,j\neq i}^{H}$. 这里 $c(s_i)$ 表示隐状态 s_i 出现时长的概率分布.

给定 M 条示教轨迹 $\boldsymbol{D}=\{\{\boldsymbol{\xi}_{n,m}\}_{n=1}^{N_m}\}_{m=1}^{M}$, 可利用 EM 对 HMM 或 HSMM 的参数进行优化[41]. 以 HMM 为例, 即

$$\boldsymbol{\theta}^*=\underset{\boldsymbol{\theta}}{\operatorname{argmax}}\,\mathcal{P}(\boldsymbol{D}|\boldsymbol{\theta}) \tag{9}$$

在通过学习训练样本获得 HMM 或 HSMM 的参数之后, 可以依据隐状态的初始概率以及状态之间的转换概率生成隐状态的序列, 同时根据这些隐状态对应的观测输出概率生成新的轨迹.

HMM 或 HSMM 的优点在于可以同时学习多种类型的轨迹, 而不需要预先对技能轨迹进行分类[42]. 然而, 和 GMM 类似, 该类方法常用于技能复现, 不易于将训练轨迹泛化到新的情形. 需要注意

的是, 当 HMM 或 HSMM 用于轨迹规划时, 通常难以产生平滑的轨迹[43]. 因此, 文献 [37] 通过加权最小二乘将多阶轨迹 (包括位置、速度和加速度) 转换为低阶的位置轨迹. 文献 [44] 将 HSMM 和模型预测控制 (Model predictive control, MPC) 相结合来获得连续的轨迹.

2.3　高斯过程 (GP)

GP (文献 [45], 第 2.2 节) 是指一系列随机变量的集合, 其中任意有限个随机变量的联合概率服从高斯分布. 特别地, 给定训练数据集合 $\{\boldsymbol{s}_n, y_n\}_{n=1}^N$, 以及假设输入 $\boldsymbol{s}$ 和其对应的观测输出 $y \in \mathbf{R}$ 之间存在某种函数关系 $y = f(\boldsymbol{s}) + \epsilon$, 其中 $\epsilon \sim \mathcal{N}(0, \sigma^2)$表示方差为 σ^2 的噪声, 那么给定新的测试输入 $\boldsymbol{s}^*$, 其对应的函数值 $f(\boldsymbol{s}^*)$ 和训练样本的输出 $\boldsymbol{y} = [y_1\ y_2\ \cdots\ y_N]^{\mathrm{T}}$ 存在如下关系:

$$\begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{s}^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} + \sigma^2 \boldsymbol{I} & \boldsymbol{k}^{*\mathrm{T}} \\ \boldsymbol{k}^* & k(\boldsymbol{s}^*, \boldsymbol{s}^*) \end{bmatrix} \right) \tag{10}$$

其中 $\boldsymbol{I}$ 是 N 维单位矩阵,

$$\boldsymbol{K} = \begin{bmatrix} k(\boldsymbol{s}_1, \boldsymbol{s}_1) & k(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & k(\boldsymbol{s}_1, \boldsymbol{s}_N) \\ k(\boldsymbol{s}_2, \boldsymbol{s}_1) & k(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & k(\boldsymbol{s}_2, \boldsymbol{s}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{s}_N, \boldsymbol{s}_1) & k(\boldsymbol{s}_N, \boldsymbol{s}_2) & \cdots & k(\boldsymbol{s}_N, \boldsymbol{s}_N) \end{bmatrix} \tag{11}$$

$$\boldsymbol{k}^* = [k(\boldsymbol{s}^*, \boldsymbol{s}_1)\ k(\boldsymbol{s}^*, \boldsymbol{s}_2)\ \cdots\ k(\boldsymbol{s}^*, \boldsymbol{s}_N)] \tag{12}$$

这里 $k(\cdot, \cdot)$ 表示核函数, 一个常见的例子是平方指数 (Squared exponential, SE) 函数 $k(\boldsymbol{s}^*, \boldsymbol{s}_n) = \exp(-\dfrac{1}{2\ell^2}||\boldsymbol{s}^* - \boldsymbol{s}_n||^2)$. 关于核函数的内容可以参考文献 [46].

根据式 (10) 中的联合概率分布和多变量高斯的条件概率分布, 可获得 $\mathcal{P}(f(\boldsymbol{s}^*)|\boldsymbol{y})$, 其均值和方差分别为 (文献 [45], 第 2.2 节):

$$\mathrm{E}(f(\boldsymbol{s}^*)) = \boldsymbol{k}^*(\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y} \tag{13}$$

$$\mathrm{D}(f(\boldsymbol{s}^*)) = k(\boldsymbol{s}^*, \boldsymbol{s}^*) - \boldsymbol{k}^*(\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{k}^{*\mathrm{T}} \tag{14}$$

式 (13) ~ (14) 仅针对训练样本中输出是一维的情形. 对于多维输出, 可以分别对每个输出变量利用 GP 进行预测, 也可以采用向量值 (Vector valued) GP 以及可分离核函数 (Separable kernels)[47].

作为典型的监督学习算法, GP 可以通过学习示教轨迹实现运动技能的复现. 对于轨迹的自适应问题, 如机器人末端从 A 点出发, 在 B 点抓取一个物体并最终将物体放置到 C 点 (这里 A, B, C 点的位置均不同于示教轨迹), 利用多变量高斯的后验概率 (Posterior) 即能够规划新的满足任务要求的轨迹. 然而, 如果利用 GP 对位置和速度分别进行预测, 则无法保证速度变量和位置变量之间的一阶微分关系. 目前基于 GP 的模仿学习文献常仅学习位置轨迹或忽略该微分约束, 事实上该问题可以利用微分 (Derivative) GP 解决[48]. 以时间输入为例, 在定义 GP 的协方差时可利用 $\mathrm{cov}\left\langle \dfrac{\mathrm{d}f(t_i)}{\mathrm{d}t_i}, f(t_j) \right\rangle = \dfrac{\mathrm{d}k(t_i, t_j)}{\mathrm{d}t_i}$ 以及 $\mathrm{cov}\left\langle \dfrac{\mathrm{d}f(t_i)}{\mathrm{d}t_i}, \dfrac{\mathrm{d}f(t_j)}{\mathrm{d}t_j} \right\rangle = \dfrac{\mathrm{d}^2 k(t_i, t_j)}{\mathrm{d}t_i \mathrm{d}t_j}$.

2.4　动态运动基元 (DMP)

DMP[2] 本质上是从示教轨迹中学习位置 $\boldsymbol{\xi}$ 和速度 $\dot{\boldsymbol{\xi}}$ 到加速度 $\ddot{\boldsymbol{\xi}}$ 的映射函数. 对于机器人系统, 假设当前时刻 t 的位置和速度 $\{\boldsymbol{\xi}_t, \dot{\boldsymbol{\xi}}_t\}$ 是可观测的, DMP 能够在线的计算期望的加速度 $\hat{\ddot{\boldsymbol{\xi}}}_t$, 由此可获得下一时刻的期望位置 (即 $\boldsymbol{\xi}_t + \delta_t \dot{\boldsymbol{\xi}}_t$) 和期望速度 (即 $\dot{\boldsymbol{\xi}}_t + \delta_t \hat{\ddot{\boldsymbol{\xi}}}_t$), 其中 δ_t 表示机器人的伺服周期. 随着时间的增加即可完成轨迹的规划任务. 同样地, DMP 可以对关节轨迹、力和力矩轨迹等进行规划.

给定一条长度为 N 的轨迹 $\{t_n, \boldsymbol{\xi}_n, \dot{\boldsymbol{\xi}}_n, \ddot{\boldsymbol{\xi}}_n\}_{n=1}^N$, DMP 使用如下模型对运动轨迹进行编码 (Encoding)[2]:

$$\tau \dot{z} = -\alpha z \tag{15}$$

$$\tau^2 \ddot{\boldsymbol{\xi}} = \boldsymbol{K}^p(\boldsymbol{g} - \boldsymbol{\xi}) - \tau \boldsymbol{K}^v \dot{\boldsymbol{\xi}} + z\boldsymbol{f}(z) \tag{16}$$

$$f_i(z) = \frac{\sum_{h=1}^{H} \varphi_h(z) w_{i,h}}{\sum_{h=1}^{H} \varphi_h(z)} (g_i - \xi_{0i}) \tag{17}$$

在式 (15) 中, $\alpha > 0$ 为常数, τ 为轨迹时长, z 表示相位变量. 该模型用来将时间信号 t 转换成 z. 在式 (16) 中, $\boldsymbol{g} \in \mathbf{R}^{\mathcal{O}}$ 表示轨迹的目标位置, $\boldsymbol{K}^p$ 和 $\boldsymbol{K}^v$ 分别表示预先设定的对角的刚度和阻尼矩阵, $\boldsymbol{f}(z) \in \mathbf{R}^{\mathcal{O}}$ 为轨迹调整项. 式 (17) 的 $f_i(z)$ 为 $\boldsymbol{f}(z)$ 的第 i 个分量的定义, 其中 $w_{i,h}$ 为加权系数, H 为拟合 $\boldsymbol{f}$ 所需要的基函数 (Basis function) 的数量, $\varphi_i(z) = \mathrm{e}^{-a_i(z-c_i)^2}$ 表示基函数, 这里 $a_i > 0$, $c_i \in [0, 1]$. g_i 和 ξ_{0i} 分别对应目标位置 $\boldsymbol{g}$ 和初始位置 $\boldsymbol{\xi}_0$ 的第 i 个分量.

在训练 DMP 时可以对式 (15) ~ (16) 进行离散化, 即利用 $\dot{z}_t = (z_{t+1} - z_t)/\delta_t$ 和 $\dot{\boldsymbol{\xi}}_t = (\boldsymbol{\xi}_{t+1} - \boldsymbol{\xi}_t)/\delta_t$, 然后通过回归算法 (如最小二乘或局部加权回归[49] (Locally weighted regression, LWR)) 估计形状参数 $\boldsymbol{W} = \{\{w_{i,h}\}_{h=1}^H\}_{i=1}^{\mathcal{O}}$. 在应用 DMP 进行泛化时,

通过调整 τ 和 $\boldsymbol{g}$ 就能够改变期望轨迹的时长 (即运动的快慢) 以及期望的目标位置.

DMP 的主要优点是可以从任意的起始点 (Start-point) 对轨迹进行规划并收敛到任意的目标点 (End-point), 而不需要其他的预处理, 如文献 [5] 需要将轨迹投影到局部坐标系中. 文献 [2] 表明, 当运动时间趋于无穷时式 (15) 中的 z 趋于零, 这时式 (16) 对应的稳定收敛点为: $\boldsymbol{\xi}=\boldsymbol{g}$ 以及 $\dot{\boldsymbol{\xi}}=\ddot{\boldsymbol{\xi}}=\boldsymbol{0}$. 然而, 在实际应用中轨迹的期望时长 τ 通常是有限的, 即当 $t=\tau$ 时 z 仍大于 0, 这时 $\boldsymbol{\xi}$ 和 $\boldsymbol{g}$ 仍会存在一定的误差.

另外, 由于 DMP 收敛时的速度为零, 导致其不适用于存在速度要求的任务 (如打乒乓球机器人需要以某期望的速度击球), 而且 DMP 无法生成经过任意中间点 (Via-point) 的轨迹. Kober 等[50] 对 DMP 进行了改进, 使其能够以非零的速度到达收敛位置, 然而仍未能处理期望中间点的问题. 除此之外, DMP 需要预设置的参数较多, 特别是基函数的选择. 为了避免基函数的问题, Fanger 等[51] 利用 GP 预测 z 对应的 $\boldsymbol{f}(z)$. 文献 [9, 52] 通过 GMM 和 GMR 预测 $\boldsymbol{f}(z)$ 在不同时刻的概率分布, 从而实现 DMP 框架下对多个示教轨迹的学习.

最后, DMP 中的参数 (即 τ, $\boldsymbol{g}$ 和 $\boldsymbol{W}$) 可以利用强化学习对其进行优化[53−56], 但需要事先根据特定的任务设计合理的成本函数. 由于强化学习采用学习和探索 (Exploitation and exploration) 的方式, 常常需要大量的迭代, 特别是当学习复杂轨迹时需要大量的基函数从而导致 $\boldsymbol{W}$ 的维度较大, 故该思路不适用于实时的技能泛化.

2.5 动态系统稳定估计 (SEDS)

SEDS[3] 利用 GMM 和 GMR 学习示教轨迹中位置 $\boldsymbol{\xi}$ 和速度 $\dot{\boldsymbol{\xi}}$ 的函数关系并通过 (非线性) 优化 GMM 的参数来获得稳定的动态系统. 给定 M 条示教轨迹 $\boldsymbol{D}=\{\{\boldsymbol{\xi}_{n,m},\dot{\boldsymbol{\xi}}_{n,m}\}_{n=1}^{N_m}\}_{m=1}^{M}$, 可以依据式 (1) 估计 $\mathcal{P}(\boldsymbol{\xi},\dot{\boldsymbol{\xi}})$, 再用式 (2) 计算 $\mathcal{P}(\dot{\boldsymbol{\xi}}|\boldsymbol{\xi})$, 以及式 (7) 估计 $\dot{\boldsymbol{\xi}}$ 对应的条件期望[2]:

$$\dot{\boldsymbol{\xi}}=\sum_{c=1}^{C}h_c(\boldsymbol{\xi})\left(\boldsymbol{\mu}_{\dot{\xi},c}+\boldsymbol{\Sigma}_{\dot{\xi}\xi,c}\boldsymbol{\Sigma}_{\xi\xi,c}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu}_{\xi,c})\right) \tag{18}$$

可进一步将式 (18) 变形为:

$$\dot{\boldsymbol{\xi}}=\sum_{c=1}^{C}h_c(\boldsymbol{\xi})\left(\boldsymbol{A}_c\boldsymbol{\xi}+\boldsymbol{b}_c\right) \tag{19}$$

其中, $\boldsymbol{A}_c=\boldsymbol{\Sigma}_{\dot{\xi}\xi,c}\boldsymbol{\Sigma}_{\xi\xi,c}^{-1}$, $\boldsymbol{b}_c=\boldsymbol{\mu}_{\dot{\xi},c}-\boldsymbol{\Sigma}_{\dot{\xi}\xi,c}\boldsymbol{\Sigma}_{\xi\xi,c}^{-1}\boldsymbol{\mu}_{\xi,c}$. 式 (19) 可以看作是 C 个由 $h_c(\boldsymbol{\xi})$ 加权的线性子系统的叠加, 而且式 (19) 中的预测模型只依赖 GMM 的参数.

为了获得稳定的系统, 文献 [3] 给出了系统稳定的充分条件, 即对于任意第 $c\in\{1,2,\cdots,C\}$ 个子系统均需要满足:

$$\boldsymbol{A}_c+\boldsymbol{A}_c^{\mathrm{T}}\prec 0 \tag{20}$$

$$\boldsymbol{A}_c\boldsymbol{\xi}^*+\boldsymbol{b}_c=\boldsymbol{0} \tag{21}$$

其中, '$\prec 0$' 表示矩阵的负定, $\boldsymbol{\xi}^*$ 为所有的子系统的收敛目标. 通过非线性优化器最大化示教轨迹的观测概率并满足上述稳定性的充分条件, 即可获得最优的 GMM 参数.

由于 SEDS 将轨迹规划问题转化成稳定的动态系统问题, 其和 DMP 一样适用于将轨迹从任意的起点泛化到任意的目标点. 然而其和 DMP 也有类似的局限性, 即无法直接处理带有速度或中间路径点要求的泛化问题. 另外, SEDS 可以学习多维度的轨迹 $\boldsymbol{\xi}$, 但是如式 (19) 所示其仅适合学习 $\boldsymbol{\xi}$ 和 $\dot{\boldsymbol{\xi}}$ 之间的映射关系, 而不适用于学习输入为时间的轨迹或输入和输出对应不同类型轨迹的情形 (如在人机交互时输入对应人的双手位置, 输出为机器人关节角度).

2.6 概率运动基元 (ProMP)

ProMP[4] 应用如下模型对示教轨迹进行拟合:

$$\begin{bmatrix}\boldsymbol{\xi}(t)\\ \dot{\boldsymbol{\xi}}(t)\end{bmatrix}=\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{w} \tag{22}$$

其中 $\boldsymbol{\Phi}(t)=[\boldsymbol{I}_{\mathcal{O}}\otimes\boldsymbol{\phi}(t)\ \boldsymbol{I}_{\mathcal{O}}\otimes\dot{\boldsymbol{\phi}}(t)]\in\mathbf{R}^{B\mathcal{O}\times 2\mathcal{O}}$, $\boldsymbol{\phi}(t)=[\varphi_1(t)\ \varphi_2(t)\ \ldots\ \varphi_B(t)]^{\mathrm{T}}$ 表示 B 维的基函数向量, $\otimes$ 为矩阵 Kronecker 乘积, $\boldsymbol{w}\in\mathbf{R}^{B\mathcal{O}}$ 为未知的轨迹参数. 注意, 如果采用 $\boldsymbol{\Phi}(t)=\boldsymbol{I}_{2\mathcal{O}}\otimes\boldsymbol{\phi}(t)$, 则无法保证预测输出中的微分关系.

给定 M 条示教轨迹, 通过动态时间规整 (Dynamic time warping, DTW) 对其进行预处理可获得长度均为 N 的轨迹, 即 $\{\{t_{n,m},\boldsymbol{\xi}_{n,m},\dot{\boldsymbol{\xi}}_{n,m}\}_{n=1}^{N}\}_{m=1}^{M}$, 然后利用最大似然估计 (Maximum likelihood estimation, MLE) 可以求取轨迹参数 $\boldsymbol{w}$ 的概率分布. 具体来说, 先用式 (22) 分别拟合不同的轨迹, 并利用最小二乘获得各个轨迹的参数 $\{\boldsymbol{w}_m\}_{m=1}^{M}$. 然后可计算 $\boldsymbol{w}\sim\mathcal{N}(\boldsymbol{\mu}_w,\boldsymbol{\Sigma}_w)$ 的概率分布[57]:

$$\boldsymbol{\mu}_w=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{w}_m \tag{23}$$

$$\boldsymbol{\Sigma}_w=\frac{1}{M}\sum_{m=1}^{M}(\boldsymbol{w}_m-\boldsymbol{\mu}_w)(\boldsymbol{w}_m-\boldsymbol{\mu}_w)^{\mathrm{T}} \tag{24}$$

[2] 将式 (2) 中的 $\boldsymbol{s}$ 和 $\boldsymbol{\xi}$ 分别用 $\boldsymbol{\xi}$ 和 $\dot{\boldsymbol{\xi}}$ 进行替换即可.

需要指出的是当 $\boldsymbol{w}$ 的维度 $B\mathcal{O}$ 大于样本数量 M 时, $\boldsymbol{\Sigma}_w$ 为奇异阵, 因此常需要引入附加的正则项, 即 $\boldsymbol{\Sigma}_w+\lambda\boldsymbol{I}$. 然而如果 λ 过小, 在应用高斯条件概率进行轨迹调整时, 常会出现数值问题. 如果 λ 过大, 正则化后的方差则会高估多条轨迹之间的方差特征.

在获得轨迹参数 $\boldsymbol{w}$ 的概率分布之后, 针对技能的复现问题, 可以直接利用 $\boldsymbol{\mu}_w$ 或从 $\mathcal{N}(\boldsymbol{\mu}_w,\boldsymbol{\Sigma}_w)$ 采样出 $\boldsymbol{w}$, 相应的复现轨迹可由式 (22) 得到. 针对轨迹的自适应问题, 可以利用条件高斯 (文献 [35], 第 2.3.1 和 2.3.3 节) 进行计算. 假设泛化的轨迹需要在特定的时刻 t^* 以期望的速度 $\dot{\boldsymbol{\xi}}_t^*$ 经过期望的位置 $\boldsymbol{\xi}_t^*$, 并且假设期望点 $\boldsymbol{\mu}_t^*=[\boldsymbol{\xi}_t^{*\mathrm{T}}\dot{\boldsymbol{\xi}}_t^{*\mathrm{T}}]^{\mathrm{T}}$ 的协方差[3] 为 $\boldsymbol{\Sigma}_t^*$, 则调整后的轨迹参数的概率分布 $\mathcal{N}(\boldsymbol{\mu}_w^*,\boldsymbol{\Sigma}_w^*)$ 为[4]:

$$\boldsymbol{\mu}_w^*=\boldsymbol{\mu}_w+\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(t^*)\boldsymbol{L}^{-1}(\boldsymbol{\mu}_t^*-\boldsymbol{\Phi}^{\mathrm{T}}(t^*)\boldsymbol{\mu}_w) \tag{25}$$

$$\boldsymbol{\Sigma}_w^*=\boldsymbol{\Sigma}_w-\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(t^*)\boldsymbol{L}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(t^*)\boldsymbol{\Sigma}_w \tag{26}$$

其中 $\boldsymbol{L}=\boldsymbol{\Phi}^{\mathrm{T}}(t^*)\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(t^*)+\boldsymbol{\Sigma}_t^*$. 最后, 可应用 $\boldsymbol{\mu}_w^*$ 或从 $\mathcal{N}(\boldsymbol{\mu}_w^*,\boldsymbol{\Sigma}_w^*)$ 采样得到的 $\boldsymbol{w}$, 根据式 (22) 生成自适应的轨迹. 该轨迹能够在期望的时刻, 在预定的方差范围内经过期望点. 对于存在多个期望点的情况, 可以依次用式 (25) 和 (26) 对 $\boldsymbol{w}$ 的概率分布进行更新.

ProMP 可以同时对位置和速度轨迹进行学习和泛化, 计算效率高, 适用于在线规划. ProMP 和 DMP 类似, 两者均用来学习时间驱动的轨迹 (即轨迹的输入为时间), 且都需要事先指定用来拟合轨迹的基函数 $\boldsymbol{\phi}(t)$. 然而, 对于高维输入的情况, 常常需要大量的基函数[4], 因此难以将 ProMP 应用于学习带有多维输入轨迹的情形. 另外, ProMP 未考虑轨迹规划中的外插问题 (即待规划轨迹从整体上偏离示教区域)[58].

2.7　核化运动基元 (KMP)

KMP[6] 从信息论的角度研究示教轨迹的模仿学习问题. 给定 M 条示教轨迹 $\{\{\boldsymbol{s}_{n,m},\boldsymbol{\xi}_{n,m}\}_{n=1}^{N_m}\}_{m=1}^{M}$, 首先利用 GMM 获得 $\mathcal{P}(\boldsymbol{s},\boldsymbol{\xi})$, 然后从 GMM 中采样[5] N 个可以表征输入空间分布特征的参考输入 $\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_N\}_{n=1}^{N}$. 应用式 (6) 计算不同参考输入 $\boldsymbol{s}_n$ 对应输出 $\hat{\boldsymbol{\xi}}_n$ 的概率分布, 即 $\hat{\boldsymbol{\xi}}_n|\boldsymbol{s}_n\sim\mathcal{N}(\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\Sigma}}_n)$, 可得到参考轨迹 $\{\boldsymbol{s}_n,\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\Sigma}}_n\}_{n=1}^{N}$. 在获得参考轨迹之后, KMP 采用如下参数化模型:

$$\boldsymbol{\xi}(\boldsymbol{s})=\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{s})\boldsymbol{w} \tag{27}$$

其中 $\boldsymbol{\Phi}(\boldsymbol{s})=\boldsymbol{I}_{\mathcal{O}}\otimes\boldsymbol{\phi}(\boldsymbol{s})\in\mathbf{R}^{B\mathcal{O}\times\mathcal{O}}$, $\boldsymbol{\phi}(\boldsymbol{s})$ 表示 B 维的基函数向量, $\boldsymbol{w}\sim\mathcal{N}(\boldsymbol{\mu}_w,\boldsymbol{\Sigma}_w)$. 这里 $\boldsymbol{\mu}_w$ 和 $\boldsymbol{\Sigma}_w$ 未知. 为了估计 $\boldsymbol{\mu}_w$ 和 $\boldsymbol{\Sigma}_w$, KMP 对式 (27) 生成轨迹的概率分布和参考轨迹的概率分布之间的 KL 散度进行最小化, 即

$$\sum_{n=1}^{N}\mathrm{KL}(\mathcal{P}_n^{para}||\mathcal{P}_n^{ref}) \tag{28}$$

其中, $\mathcal{P}_n^{para}=\mathcal{N}\big(\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{s}_n)\boldsymbol{\mu}_w,\ \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{s}_n)\boldsymbol{\Sigma}_w\boldsymbol{\Phi}(\boldsymbol{s}_n)\big)$, $\mathcal{P}_n^{ref}=\mathcal{N}\big(\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\Sigma}}_n\big)$. 对该目标函数进行分解, 利用向量和矩阵求导以及核技巧可获得任意输入 $\boldsymbol{s}^*$ 对应轨迹 $\boldsymbol{\xi}(\boldsymbol{s}^*)$ 的均值和协方差[6]:

$$\mathrm{E}(\boldsymbol{\xi}(\boldsymbol{s}^*))=\ \boldsymbol{k}^*(\boldsymbol{K}+\lambda_1\boldsymbol{\Sigma})^{-1}\boldsymbol{\mu} \tag{29}$$

$$\mathrm{D}(\boldsymbol{\xi}(\boldsymbol{s}^*))=\frac{N}{\lambda_2}\left(\boldsymbol{k}(\boldsymbol{s}^*,\boldsymbol{s}^*)-\boldsymbol{k}^*(\boldsymbol{K}+\lambda_2\boldsymbol{\Sigma})^{-1}\boldsymbol{k}^{*\mathrm{T}}\right) \tag{30}$$

其中 $\lambda_1>0$ 和 $\lambda_2>0$ 为正则化系数. $\boldsymbol{k}^*\in\mathbf{R}^{\mathcal{O}\times N\mathcal{O}}$ 为 $1\times N$ 的分块矩阵, 其第 i 列为 $k(\boldsymbol{s}^*,\boldsymbol{s}_i)\boldsymbol{I}_{\mathcal{O}}$. $\boldsymbol{K}\in\mathbf{R}^{N\mathcal{O}\times N\mathcal{O}}$ 为 $N\times N$ 的分块矩阵, 其第 i 行第 j 列为 $k(\boldsymbol{s}_i,\boldsymbol{s}_j)\boldsymbol{I}_{\mathcal{O}}$. $\boldsymbol{\mu}=[\hat{\boldsymbol{\mu}}_1^{\mathrm{T}}\ \hat{\boldsymbol{\mu}}_2^{\mathrm{T}}\ \cdots\ \hat{\boldsymbol{\mu}}_N^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{\Sigma}=\mathrm{blockdiag}\{\hat{\boldsymbol{\Sigma}}_1,\hat{\boldsymbol{\Sigma}}_2,\cdots,\hat{\boldsymbol{\Sigma}}_N\}$.

对于技能的复现问题, 可以直接应用式 (29) 进行轨迹规划. 对于经过期望路径点的自适应问题, 如记 M 个期望点的集合为 $\{\bar{\boldsymbol{s}}_m,\bar{\boldsymbol{\mu}}_m,\bar{\boldsymbol{\Sigma}}_m\}_{m=1}^{M}$, 其中 $\bar{\boldsymbol{s}}_m$, $\bar{\boldsymbol{\mu}}_m$ 和 $\bar{\boldsymbol{\Sigma}}_m$ 分别为第 m 个期望点的输入、输出的期望和协方差, 可直接将该期望点集合和参考轨迹 $\{\boldsymbol{s}_n,\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\Sigma}}_n\}_{n=1}^{N}$ 串联成长度为 $N+M$ 的轨迹[6], 这时应用式 (29) 学习新的扩展轨迹即可获得经过所有期望路径点的自适应的轨迹.

除了学习带有多维输入的示教轨迹, 作为 KMP 的一个特殊情况, KMP 也能够学习时间驱动的轨迹, 并同时对位置和速度进行泛化. 和多维输入情况相比, 只需要用式 (22) 替换式 (27), 利用 $\dot{\boldsymbol{\varphi}}(t)\approx\dfrac{\boldsymbol{\varphi}(t+\delta_t)-\boldsymbol{\varphi}(t)}{\delta_t}$ 以及核函数 $\boldsymbol{\phi}^{\mathrm{T}}(t_i)\boldsymbol{\phi}(t_j)=k(t_i,t_j)$ 即可. 另外, 文献 [6] 引入了任务参数化的处理方式, 使得 KMP 能够在远离示教的区域处理外插问题. 然而, KMP 未考虑轨迹中的动态问题, 无法确保轨迹的收敛性. 从计算效率上看, KMP 和 GP 的计算复杂度为 $\mathrm{O}(N^3)$, 当参考轨迹长度特别大时, 两者均不适用于在线的自适应问题. 对于这种情形, 可以利用近似方法提高学习效率, 如投影过程近似

[3] 该协方差可以控制自适应轨迹经过期望点 $\boldsymbol{\mu}_t^*$ 的误差: $\boldsymbol{\Sigma}_t^*$ 越小则误差越小, 反之则误差变大.

[4] 根据文献 ([35], 第 3.6 节), 固定基函数的数量常随输入变量维度的增加呈指数级增加.

[5] 关于从 GMM 中采样的方法可以参考文献 [59].

[6] 对于期望点输入和参考轨迹存在重叠的情况, 可参考文献 [6] 中的轨迹更新策略.

(Projected process approximation) 等 (文献 [45], 第 8.3 节).

值得一提的是, 如果将式 (29) 中的所有参考轨迹的方差 $\hat{\boldsymbol{\Sigma}}_n$ 替换成 $\boldsymbol{I}_{\mathcal{O}}$, 则 KMP 的均值退化成 GP 的均值. 如果将 $\hat{\boldsymbol{\Sigma}}_n$ 替换成 $c_n\boldsymbol{I}_{\mathcal{O}}$, 这里 $c_n > 0$ 为常量, 则式 (29) 等价于异方差高斯过程 (Heteroscedastic Gaussian processes, HGP)[60−61] 的均值预测. 和 GP、HGP 最大的区别是 KMP 在预测中显性的引入样本轨迹的方差, 并且可以预测多输出变量对应的协方差.

表 1 对本节所讨论的方法进行了总结 (部分内容来自文献 [6]), 包括 i) 技能复现; ii) 学习多条示教轨迹的概率分布, 包括期望和方差; iii) 将示教轨迹调整到经过任意的中间路径点 (位置和速度); iv) 将示教轨迹泛化到任意的目标点 (位置和速度); v) 整体偏离示教区域的泛化, 即外插; vi) 轨迹随时间的收敛性; vii) 学习带有时间输入的示教轨迹; viii) 学习带有多维动态输入的示教轨迹.

3 模仿学习中的其他若干关键问题

在模仿学习的基本问题之外, 本节将结合第 2 节中的方法对其他若干关键问题及相关文献进行综述. 需要说明的是, 本节中所讨论的问题尽管在研究内容上存在差异, 但这些方法在实质上均与轨迹规划相关.

3.1 姿态和刚度矩阵的学习问题

3.1.1 姿态的学习

表 1 中的方法可以学习无约束的轨迹, 如机器人末端位置和速度、关节位置和速度、力和力矩轨迹等. 然而在学习机器人末端姿态时, 需要考虑相应的姿态约束, 如四元数 (Quaternion) $\boldsymbol{q} \in \mathbf{S}^3$ 需要满足 $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{q} = 1$, 旋转矩阵 (Rotation matrix) $\boldsymbol{R}$ 则需要为正交矩阵, 即 $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{I}$. 这里主要依据文献 [34] 并以四元数姿态为例进行讨论.

对于学习四元数姿态的问题, 如果在 $\mathbf{R}^3$ 空间上对姿态的四个元素分别进行学习 (如 Pastor 等[62] 利用 DMP, Silverio 等[63] 采用基于 GMM[5] 的方法) 则生成的姿态轨迹无法满足单位范数的要求. 为了满足姿态约束, Ude 等[64] 和 Abu-Dakka 等[65] 均利用四元数的几何特性对 DMP 进行扩展, 其主要思路是将当前姿态和目标姿态的距离转化到 $\mathbf{R}^3$ 空间, 然后用变换后的距离替换式 (16) 中的位置距离 $\boldsymbol{g} - \boldsymbol{\xi}$. Ravichandar 等[66] 采用类似的处理方法将 SEDS[3] 应用到姿态学习中, 其中自治系统对应的输入为角速度和转换到 $\mathbf{R}^3$ 的姿态距离, 输出为角加速度. 这类基于动态模型的方法保留了原有方法的优点和局限性, 如可以朝着任意的目标姿态进行泛化以及具有收敛性, 然而其无法处理带有角速度或中间路径点要求的问题.

Zeestraten 等[67] 从黎曼几何 (Riemannian manifold) 的角度研究多条姿态轨迹的概率分布, 其主要依赖两个映射: 对数映射 (Logarithmic map) 和指数映射 (Exponential map). 前者可以将姿态投影到相应的切空间 (Tangent space), 后者被用于从切空间中恢复姿态. 由于在概率建模时存在不同的切空间, 文献 [67] 利用平行迁移 (Parallel transport) 实现不同切空间中投影的迁移. 另外, 文献 [67] 引入了任务参数化[5] 的技巧, 因此可应用于目标姿态的自适应问题. 然而, 文献 [67] 未考虑与角速度或中间路径点相关的泛化问题.

Huang 等[30, 34] 采用文献 [64] 中的空间变换方法, 将 KMP 扩展到姿态学习中. 该方法除了可以处理姿态的中间路径点和目标点问题 (包括姿态和角速度), 也考虑了角加速度或角加加速度最小化的问题. 另外, 文献 [34] 也适用于学习以及泛化带有

表 1 几种主要模仿学习方法的对比

Table 1 Comparison among the state-of-the-art approaches in imitation learning

	技能复现	多轨迹概率	中间点		目标点		外插	收敛性	时间输入	多维输入
			位置	速度	位置	速度				
GMM[35]	√	√	—	—	—	—	—	—	√	√
HMM/HSMM[40]	√	√	—	—	—	—	—	—	√	√
GP[45]	√	—	√	√	√	√	—	—	√	√
DMP[2]	√	—	—	—	√	—	√	√	√	—
SEDS[3]	√	√	—	—	√	—	√	√	—	√
ProMP[4]	√	√	√	√	√	√	—	—	√	—
KMP[6]	√	√	√	√	√	√	√	—	√	√
TP-GMM[5]	√	√	—	—	√	—	√	—	√	√

多维输入的姿态轨迹. 然而, 文献 [34] 的一个主要局限性在于其假设多条示教轨迹中的姿态在同一时刻应处在 $\mathbf{S}^3$ 中的同一个半球面, 因此不适用于多条姿态轨迹分布在不同半球面的情形.

上述所有方法的学习对象均为完整的姿态轨迹, 而不涉及姿态轨迹的分割问题. 与之不同的是, Saveriano 等[68] 提出通过学习多个 DMP 来处理中间路径点的问题. 以一个中间点为例, 其思路为先应用第一个 DMP 生成一条从起始姿态到中间姿态的轨迹, 而后用第二个 DMP 生成从中间姿态到目标姿态的轨迹. 该方法的主要缺点是需要根据中间路径点的数量对示教轨迹进行分割并分别用来训练不同的 DMP, 因此难以扩展到带有任意多个 (如大于 1) 中间路径点的问题. 另外对轨迹采取分段泛化的方式无法确保组合后的轨迹其在整体形状上与示教轨迹的相似性.

表 2 对本节中姿态学习的方法进行了总结 (主要内容来自文献 [34]), 其中 "单位范数"是指生成的轨迹满足 $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{q}=1$, "中间姿态"中的 "单个基元"是指单独的运动基元能够实现中间姿态的泛化问题.

3.1.2　刚度矩阵的学习

对于刚度矩阵的学习, 可以采用和文献 [67] 类似的基于黎曼几何的方法, 其主要步骤包括刚度矩阵和其切空间之间的映射以及利用迁移函数实现不同切空间中投影的迁移. Abu-Dakka 等[69] 将该思路推广到 DMP 框架中, 实现了 DMP 对刚度矩阵的学习, 后又将其与 KMP 进行结合[70], 实现了刚度矩阵朝着任意期望刚度状态的泛化. 需要说明的是黎曼几何方法可以学习任意的对称正定矩阵 (Symmetric positive definite, SPD), 如刚度 (Stiffness) 和阻尼 (Damping) 矩阵. Calinon[71] 对基于黎曼几何的模仿学习方法进行了总结.

学习刚度矩阵 $\boldsymbol{K}$ 的另一种方法是采用矩阵的 Cholesky 分解[72], 即 $\boldsymbol{K}=\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}$, 将 $\boldsymbol{L}$ 中的元素串成向量 $\boldsymbol{l}$ 后, 可直接对该向量进行概率建模和学习, 最后利用新生成的 $\boldsymbol{l}$ 可恢复 $\boldsymbol{L}$ 并计算出 $\boldsymbol{K}$. Wu 等[73] 在学习人体手臂末端的刚度时采用该矩阵分解的方法.

3.2　引入任务或环境变量的模仿学习问题

在模仿学习中常考虑外在的附加变量以提高机器人的学习能力, 包括任务变量 (如被抓物体的位置)、环境变量 (如障碍物的尺寸和位置) 和轨迹类型变量等. 以打乒乓球机器人为例, 可以将来球状态当作任务变量, 据此选择恰当的击打动作.

Forte 等[74] 研究了从任务变量预测 DMP 参数的问题, 其首先收集不同任务变量 $\boldsymbol{s}$ 下的运动轨迹, 然后分别提取每个运动轨迹对应的 DMP 参数, 包括目标位置 $\boldsymbol{g}$、运动时长 τ 和基函数加权系数 $\boldsymbol{W}$. 在收集足够的训练样本之后, 给定新的 $\boldsymbol{s}^*$ 应用 GP 预测其对应的 DMP 的参数 $\{\boldsymbol{g}^*,\tau^*,\boldsymbol{W}^*\}$. 最后由式 (15)~(17) 生成任务变量 $\boldsymbol{s}^*$ 条件下的轨迹. 类似地, Kramberger 等[75] 利用 LWR 对 DMP 的模型参数进行预测, 并将其应用于末端位置和姿态的学习之中. 和文献 [74–75] 不同, 文献 [76] 和 [31] 在 DMP 的轨迹调整项 $f_i(z)$ (即式 (17)) 中分别显性地引入任务变量和表示轨迹类型 (Style) 的变量 $\boldsymbol{s}$. 这时 $f_i(z)$ 变成 $f_i(z,\boldsymbol{s})$. Colome 等[13] 将 ProMP 中的参数 $\boldsymbol{w}$ 降维成 $\tilde{\boldsymbol{w}}$, 然后用 GMM 和 GMR 预测[7] 新的 $\boldsymbol{s}^*$ 对应的 $\tilde{\boldsymbol{w}}^*$, 继而用其恢复 $\boldsymbol{w}$ 来生成新的轨迹 (利用式 (22)). 上述方法的主要不足在于需要充分多的训练样本, 在小样本情况下难以进行较大范围的泛化.

Calinon 等[5, 37] 对 GMM 进行扩展, 提出了 TP-

[7] 在预测之前需要获得足够多的训练样本对 $\{\boldsymbol{s},\tilde{\boldsymbol{w}}\}$.

表 2　几种主要姿态学习方法的对比

Table 2　Comparison among the state-of-the-art approaches in orientation learning

	单位范数	多轨迹概率	中间姿态			目标姿态		收敛性	时间输入	多维输入
			单个基元	姿态	角速度	姿态	角速度			
Pastor 等[62]	—	—	—	—	—	√	—	√	√	—
Silverio 等[63]	—	√	—	—	—	√	—	—	√	√
Ude 等[64]	√	—	—	—	—	√	—	√	√	—
Abu-Dakka 等[65]	√	—	—	—	—	√	—	√	√	—
Ravichandar 等[66]	√	√	—	—	—	√	—	√	—	√
Zeestraten 等[67]	√	√	—	—	—	√	—	—	√	√
Huang 等[34]	√	√	√	√	√	√	√	—	√	√
Saveriano 等[68]	√	—	—	√	√	√	√	√	√	—

GMM, 其核心是针对不同的任务参数设计恰当的局部坐标系, 然后将示教轨迹投影到各个局部坐标系中用来学习其相对的运动特征. 如抓取任务, 这里以一个局部坐标系为例, 可以将局部坐标系设置在目标物体上, 从而能够学习机器人末端和物体之间相对距离随时间变化的特征. 当抓取其他位置的新物体时, 可将上述得到的相对距离看作是机器人和新物体之间的距离, 最后将该相对距离转换到机器人的坐标系中获得绝对位置. Silverio 等[63] 将 TP-GMM 推广到四元数姿态的学习中, 然而该方法未考虑姿态的单位范数约束. TP-GMM 中任务参数化的处理方法也被应用于文献 [6, 67] 之中.

TP-GMM 存在的一个主要的问题是: 难以事先根据机器人的任务指定最优的局部坐标系, 如根据抓取物体的位置可知局部坐标系的原点, 然而该坐标系的最优姿态是未知的. 因此, 文献 [77] 应用强化学习对局部坐标系进行优化 (包括旋转和平移), 且证明了对低维度坐标系参数的优化可以转换为对于高维度 GMM 参数的优化.

TP-GMM 可以通过学习少量的样本实现较大范围的泛化, 然而其一般仅用于学习机器人任务空间的轨迹, 难以扩展到机器人关节空间中. 另外, 对不同局部坐标系中的轨迹进行高斯相乘 (Gaussian product) 的处理方式无法保证生成轨迹的平滑性 (特别是位置轨迹对应的速度) 以及泛化精度 (即和期望的目标位置常常存在一定的误差). 表 1 对 TP-GMM的特征进行了总结.

3.3 轨迹分割、运动基元串联和叠加问题

3.3.1 运动基元的提取和串联

轨迹分割 (Segmentation) 问题是指从一个完整的轨迹中提取出一系列的基本运动单元, 也称运动基元 (Movement primitive, MP), 所获得的 MP 通过恰当的串联 (Sequence) 可以用于技能的复现和泛化. 以机器人打开冰箱取牛奶为例, 一个完整的动作包括机器人打开冰箱、抓取牛奶以及关门等对应不同子任务的动作, 其中每一个动作或子任务实质上对应一个 MP. 从该完整动作中提取出的 MP 经过合理的串联和泛化即可应用到类似的序列任务的场景中.

针对序列任务轨迹的分割, 近年来被广泛采用的一种方法是 HMM. Kulic 等[78] 应用 HMM 对示教轨迹进行分割、聚类以及 MP 的建模, 然后通过构建 MP 之间的概率转移图实现不同 MP 之间的转换, 最终形成由多个 MP 串联而成的轨迹. Manschitz 等[79] 假设所有的 MP 均由具有收敛特性的二阶动态系统表征, 其首先通过轨迹中的运动特征[80] (如速度的停顿、接触力的出现和消失等作为轨迹分割点) 对轨迹进行初步分割[8], 后在应用 HMM 提取 MP 时 (分割后的轨迹片段对应观测值, 隐状态对应 MP), 利用有向正态分布 (Directional normal distribution, DND) 对隐状态的输出观测概率分布进行建模并依据 BIC 选择最优的隐状态 (即 MP) 的数量, 其中 DND 同时考虑了轨迹的位置和速度向量, 因此可以将 MP 的收敛假设和轨迹片段的聚类相结合.

和文献 [79] 类似, Medina 等[81] 在考虑多 MP 序列问题时, 也假设了 MP 的收敛特性. 两者的主要区别是文献 [81] 中 HMM 隐状态的输出观测为变参数的动态系统. 另外, 文献 [81] 显性地引入判断 MP 终止的二进制变量, 而文献 [79] 则采用分类器对 MP 的转换进行预测.

目前针对序列任务学习的方法主要侧重于对 MP 的提取和串联, 未充分研究序列任务中单独 MP 的泛化 (如 MP 的形状和运动时长等). 文献 [79] 通过在目标物体上定义局部坐标系来学习机器人和物体之间的相对位置, 可以实现一定程度的泛化, 却难以应用于涉及运动速度、时长和轨迹形状等要求的场景. 除了 HMM, 其他的轨迹分割方法还包括如基于 MP 库匹配的方法[82] 和 GMM[83] 等.

与从序列任务的轨迹中提取 MP 不同, Pastor 等[62] 研究在给定多个 MP 的情况下对 MP 进行串联的衔接点平滑问题, 其中 MP 为 DMP. Stulp 等[84] 应用强化学习对多个串联 DMP 的参数进行优化, 并提出了分层定义误差函数的方法, 其中对于任意一个 DMP, 其形状参数 $\boldsymbol{W}$ 的误差函数仅由该 DMP 生成的轨迹决定, 而其目标点 $\boldsymbol{g}$ 的误差函数则由该 DMP 以及后续 DMP 共同决定. Daniel 等[85] 采用分层强化学习的方法对序列任务中 MP 的顺序以及各个 MP 的参数进行优化. 与文献 [84] 相比, 文献 [85] 不需要事先指定 MP 的顺序, 然而两者均需要指定 MP 的数量以及定义合理的奖励函数. 需要强调的是, 当机器人任务需要多个 MP 串联时基于强化学习的方法通常需要大量的训练, 特别是在未指定 MP 顺序以及 MP 作用下任务状态转移概率未知的情况.

3.3.2 运动基元的叠加

在对 MP 进行串联之外, 还可以对多个 MP 进行叠加 (Superposition), 如 ProMP[4] 和 KMP[6] 通过高斯概率的特性直接对 MP 进行叠加. Duan

[8] 分割后的轨迹片段一般不等同于 MP, 常常不同的轨迹片段可能对应相同的 MP, 因此需要对轨迹片段进行聚类.

等[86] 针对不同的任务轨迹设计相应的激活函数, 该函数实质上是通过调整不同轨迹的方差来实现多条轨迹间的切换. Silverio 等[87] 针对关节空间轨迹、任务空间轨迹和末端交互力分别设计力矩控制器, 其中不同的控制器可以看作是表示不同任务的 MP, 最后通过高斯乘积可以将不同的控制器合并成一个最终的力矩控制器. 实际上, 对于第 3.2 节中讨论的 TP-GMM, 如果将不同局部坐标系内的轨迹分布看作是 MP, 也可以将其理解为多个 MP 的叠加.

3.4 协同和不确定性预测的问题

3.4.1 多维轨迹的协同问题

模仿学习中的一个重要特点是学习多维轨迹的协同 (Synergy), 又称作协调 (Coordination). 以机器人和人类握手为例[9], 一个自然的握手动作主要依赖胳膊的肘关节和腕关节, 并适当地调动其他的关节. 如果对机器人手臂的关节分别进行轨迹规划是可以实现末端的握手动作, 然而整个机器人手臂在握手过程中的姿势可能是不恰当的, 特别是当握手的位置和频率发生改变时, 不同的关节需要在协同的情况下进行调整. 另一个例子是两个机械手臂的协同作业[63], 当一只手臂受到干扰时另一只手臂也应该产生相应的调整, 而非独立的对两只机械手臂进行轨迹规划.

对于多维轨迹的协同问题, 可以采用对其概率分布进行建模的方法获得轨迹中的协方差, 如第 2 节中基于概率的方法[9] GMM, HMM/HSMM, ProMP 和 KMP. 该协方差即可表征轨迹中的协同关系. 同时, 协方差也包含轨迹中不同维度的方差信息 (可以理解为多条轨迹之间的变化幅度). 以两维变量的高斯分布为例, 其均值为 2×1 的向量而协方差为 2×2 的矩阵, 如果协方差矩阵的非对角元素均为 0, 则表明两个变量是独立的, 否则变量间存在协同关系, 注意这里协方差矩阵对角线上的元素分别表示两个变量的方差. 另外, 在学习多个示教轨迹时, 常直接将轨迹点对应的协方差矩阵当做判断其重要性的一个依据, 即轨迹点的协方差和其重要性相反, 如文献 [7] 利用协方差计算轨迹之间的相似度.

3.4.2 不确定性预测的问题

不确定性 (Uncertainty) 是用来度量模仿学习生成轨迹的可信度. 以文献 [21] 中人机协同的任务为例, 其中操作者的手部位置是机器人运动的控制输入. 当人类在示教区域内时, 依据人的手部位置而预测得到的期望的机器人轨迹是可信的, 因此该轨迹的不确定性较低. 当人类远离示教的工作区域时, 其对应的预测轨迹是不可信的, 因此该轨迹的不确定性较高. 对于不确定性的预测, 可以应用 GP 和 KMP.

另外, 模仿学习中还存在一些能够同时预测轨迹协方差和不确定性的方法. 这类方法同时考虑如下两种情况: i) 当输入在示教区域内时, 预测的协方差能够对应示教轨迹之间的关联和变化程度; ii) 当输入远离示教区域时, 预测的协方差[10] 则对应预测轨迹的不确定性. Schneider 等[88] 在 HGP[61] 的框架下通过优化不同输入对应的噪声方差实现对轨迹方差的学习, 同时该方法也可以提供不确定性的预测. 然而文献 [88] 未考虑多维轨迹之间的协同问题. Umlauft 等[89] 利用多个 GP 预测的均值构建 Wishart 过程[90] 从而实现对协方差和不确定性的预测, 其中涉及的所有 GP 的参数以及其他参数可通过数值优化求解 MLE 获得. Silverio 等[21] 证明了 KMP 也可以同时对协方差和不确定性进行预测.

值得一提的是, 文献 [89] 中控制机器人轨迹跟踪的刚度矩阵是根据协方差来定义的. 在上述的两类情况 i) 和 ii) 中, 只有当输入在训练区域内并且当示教轨迹的协方差小时, 预测输出的协方差才会很小; 否则, 预测的协方差则很大. 因此, 文献 [89] 将刚度矩阵的特征值和预测协方差的特征值在大小上设置成反比关系. 文献 [21] 将 KMP 预测的协方差和线性二次型调节器 (Linear quadratic regulator, LQR) 相结合, 其中预测协方差的逆矩阵被当作 LQR 中跟踪误差的加权矩阵, 实现了变刚度和变阻尼的控制. 和文献 [21] 类似的将协方差与控制器相结合的工作[11] 还有文献 [5, 44, 91−92], 然而文献 [5, 44, 91−92] 中的协方差只针对情况 i) 而不包括不确定性的预测.

3.5 混合空间下的模仿学习问题

混合空间下的模仿学习是指机器人同时在任务空间和关节空间进行模仿学习, 其可以应用于需要同时考虑末端任务和关节姿态的场景. 以机器人在黑板上进行书写为例, 机器人末端的轨迹对完成书写任务是至关重要的, 但同时机器人的关节轨迹可以确保机器人在书写过程中的姿势是自然的、合理的.

与单空间 (任务或关节空间) 的模仿学习相比, 混合模仿学习需要考虑机器人关节轨迹和末端轨迹

[9] 向量值 GP 通过恰当的可分离核函数可以表征多维轨迹之间的协同关系, 然而其未考虑轨迹本身的方差, 故这里未将其包括在内.

[10] 这里使用 "协方差"是为了表明 i) 和 ii) 使用相同的预测模型.

[11] 这些工作中对应的控制器被称作最小干涉控制 (Minimal intervention control).

之间的正向运动学 (Forward kinematics) 约束. 文献 [93] 分别对示教的末端位置轨迹 $\boldsymbol{\xi}^p$ 和关节角度轨迹 $\boldsymbol{\xi}^q$ 用 GMM 进行建模, 后用 GMR 获得两种轨迹随时间变化的概率分布, 即 $\mathcal{P}(\boldsymbol{\xi}_t^p)$ 和 $\mathcal{P}(\boldsymbol{\xi}_t^q)$. 通过基于雅克比 (Jacobian) 矩阵的逆运动学 (Inverse kinematics), 将末端轨迹的概率分布 $\mathcal{P}(\boldsymbol{\xi}_t^p)$ 转换到关节空间, 得到 $\mathcal{P}(\tilde{\boldsymbol{\xi}}^q)$, 最后将 $\mathcal{P}(\tilde{\boldsymbol{\xi}}^q)$ 和 $\mathcal{P}(\boldsymbol{\xi}_t^q)$ 进行高斯相乘即可获得最终用于机器人控制的关节轨迹. 注意在对任务空间轨迹进行概率建模时, 文献 [93] 将任务空间的轨迹转化成相对于物体的相对距离轨迹, 该处理方式和 TP-GMM 中的任务参数法方法在实质上是相同的. Schneider 等[88] 采用同样的方法处理混合空间的学习问题, 区别是将文献 [93] 中的 GMM 和 GMR 替换成 HGP 方法[61].

除了对两个空间中的末端位置和关节角度轨迹进行学习, Calinon 等[94] 研究了双空间中速度轨迹 (即 $\dot{\boldsymbol{\xi}}^p$ 和 $\dot{\boldsymbol{\xi}}^q$) 的模仿学习, 其在统一双空间速度轨迹时和文献 [93] 类似, 亦采用雅克比矩阵的逆矩阵将任务空间的速度转换到关节空间. Paraschos 等[95] 采用 ProMP 对双空间中的加速度轨迹 (即 $\ddot{\boldsymbol{\xi}}^p$ 和 $\ddot{\boldsymbol{\xi}}^q$) 进行规划, 然后将关节加速度的概率 $\mathcal{P}(\ddot{\boldsymbol{\xi}}^q)$ 当作先验概率 (Prior), 并利用雅克比矩阵得到似然概率 $\mathcal{P}(\ddot{\boldsymbol{\xi}}^p|\ddot{\boldsymbol{\xi}}^q)$, 最后将 ProMP 生成的 $\ddot{\boldsymbol{\xi}}^p$ 当作观测值并应用条件高斯获得关节加速度的后验概率 $\mathcal{P}(\ddot{\boldsymbol{\xi}}^q|\ddot{\boldsymbol{\xi}}^p)$.

上述方法仅考虑任务空间的泛化问题, 忽视了关节轨迹的调整. 如文献 [9] 指出, 当泛化后机器人的末端轨迹远离示教区域时, 直接应用示教的关节轨迹可能是不合理的. 因此, 文献 [9] 在 DMP 的框架下研究了任务空间和关节空间同时泛化的问题, 其主要通过优化机器人雅克比矩阵对应的零空间 (Null space) 运动获得关节的最优目标位置, 并最小化泛化后关节轨迹和末端轨迹的不一致性. 该方法继承了 DMP 的局限性, 无法处理带有速度或中间路径点要求的问题.

3.6 带有任务或环境约束的模仿学习

3.6.1 带有约束的运动基元

在实际机器人系统中经常存在各种各样的约束, 如机器人关节角度、力矩和末端运动范围的限制以及避障等. 在应用模仿学习进行轨迹规划时需要将限制运动的约束因素考虑进去.

针对避障问题, Ijspeert 等[2] 提出在 DMP 的动态模型中 (即式 (16)) 增加修正量的方法, 该修正量[96] 可根据机器人和障碍物之间的距离以及机器人的速度对机器人的期望加速度进行实时调整. 增加修正量的方法也被文献 [97−98] 所采用. 文献 [9] 利用强化学习对 TP-GMM 中的局部坐标系进行优化从而实现避障. 然而文献 [2, 9, 97−98] 仅适用于机器人末端的避障问题, 未考虑机器人关节和障碍物的碰撞问题. 文献 [99] 利用 DMP 在任务空间进行规划获得期望的末端速度, 后采用文献 [100] 中调整雅克比矩阵零空间轨迹的方法实现关节空间的避障.

Shyam 等[101] 采用和文献 [102] (即 Covariant hamiltonian optimization for motion planning, CHOMP) 相同的避障函数, 利用梯度下降 (Gradient descent) 的方法对 ProMP 的参数进行迭代优化, 其中在计算避障函数时将机器人关节之间的连杆用一系列的球体 (Body point) 表示, 然后评估这些球体到障碍物的距离[102]. 因此文献 [101] 可以处理关节空间避障的问题. 注意文献 [102] 对轨迹优化时直接将轨迹当做一个未知的函数, 采用泛函梯度 (Functional gradient) 的方法计算避障函数对轨迹函数的导数[12], 而文献 [101] 中的梯度为避障函数对轨迹参数的导数, 故其利用求导的链式法则加入轨迹函数对轨迹参数的导数. 该方法的局限性是在针对避障的优化后, 新的轨迹参数可能无法严格满足优化前的泛化要求.

Huang 等[103] 在 KMP 的框架下研究带有线性约束的模仿学习问题, 该方法可以处理任意关于位置和速度的线性等式和不等式约束 (如平面约束、关节角度限制以及机器人末端运动范围的约束等), 并且能够在满足约束的情况下对轨迹进行泛化, 如在期望的时刻以期望的速度经过期望的位置. 然而该方法未考虑非线性约束. Saveriano 等[104] 将轨迹的不等式约束当作零障碍函数 (Zeroing barrier function), 通过设计恰当的控制输入使得一阶动态系统生成的轨迹满足约束条件. 文献 [103−104] 均未考虑避障问题. 值得一提的是文献 [105] 近来对 KMP 进行了扩展, 该方法能够处理带有线性和非线性、等式和不等式约束的模仿学习问题, 且可以同时考虑机器人关节的避障问题.

3.6.2 带有约束的轨迹序列的优化

如果将轨迹看作 N 个离散点 $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_N\}$ (如等时间间隔的关节轨迹点), 可以直接对由离散点串联而成的向量 $\boldsymbol{\zeta} = [\boldsymbol{q}_1^{\mathrm{T}}\ \boldsymbol{q}_2^{\mathrm{T}}\ \cdots\ \boldsymbol{q}_N^{\mathrm{T}}]^{\mathrm{T}}$ 进行优化. Osa 等[106] 利用泛函梯度[102] 对机器人的关节轨迹进行优化, 同时考虑关节避障问题以及关节轨迹对应的末端轨迹与示教末端轨迹的匹配问题, 并且给出通过

[12] 利用泛函梯度得到的导数为函数, 该导数用来对函数本身进行优化.

条件概率对示教的末端轨迹进行泛化的方法. 该文中末端轨迹的泛化精度依赖示教轨迹和环境变量组成的样本对的数量. 另外, 如文献 [107] 指出, 在对离散轨迹进行基于梯度下降的迭代时, 通常需要选择很小的步长来确保迭代过程中轨迹的平滑性 (Smoothness), 因此会增加迭代的次数.

Rana 等[108] 假设轨迹是由时变的随机微分方程生成, 然后可获得由高斯分布表示的微分方程的解, 即 $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{\mu}_{\zeta}, \boldsymbol{\Sigma}_{\zeta})$, 通过将避障以及期望起始点对应的具有最小二乘形式的目标函数 $f(\boldsymbol{\zeta})$ 与其进行合并, 得到非线性优化目标函数 $(\boldsymbol{\zeta}-\boldsymbol{\mu}_{\zeta})^{\mathrm{T}}\boldsymbol{\Sigma}_{\zeta}^{-1}(\boldsymbol{\zeta}-\boldsymbol{\mu}_{\zeta})+f(\boldsymbol{\zeta})$. 文献 [108] 未考虑轨迹的平滑性问题, 而且难以确保轨迹在优化过程中位置和速度的微分关系.

Koert 等[109] 对于机器人末端避障的问题先通过强化学习获得无碰撞 (Collision-free) 轨迹的概率分布, 然后将该概率分布用来训练 ProMP, 继而实现避障和轨迹泛化. 该方法的主要局限性是当障碍物位置发生变化时需要重新应用强化学习获得新的无碰撞轨迹的概率分布, 不适用于需要快速规划的场合. Ye 等[110] 结合模仿学习和基于采样的方法, 其将模仿学习生成轨迹当作参考轨迹, 在障碍物附近利用采样生成的无碰撞的位置点构建路径图 (Graph), 最后用 Dijkstra 算法寻找最优的可行路径. 该方法可以有效地实现关节空间的避障, 其局限性在于未能考虑轨迹在避障时的平滑性且难以扩展到带有速度要求的问题.

文献 [111−112] 采用逆最优控制 (Inverse optimal control, IOC) 的思路, 先优化示教轨迹对应的成本函数 (Cost function) 的参数, 后根据该函数采用受限优化技术对整个轨迹序列 $\boldsymbol{\zeta}$ 进行优化, 其中文献 [111] 利用逆 KKT (Karush-Kuhn-Tucker) 方法而文献 [112] 则利用协方差矩阵自适应进化策略[113] (即 Covariance matrix adaptation evolution strategy, CMA-ES) 对成本函数的参数进行优化. 这类方法可以考虑复杂的轨迹约束, 然而难以对轨迹进行实时的调整且不易于确保轨迹的平滑性, 特别是轨迹对应的高阶微分轨迹.

3.7 人机交互中的模仿学习问题

当模仿学习用于人机交互 (Human-robot interaction) 时需要考虑人类和机器人之间的时间同步 (Synchronization) 问题. 以人类和机器人协同搬运物品为例, 机器人需要根据人的状态 (如位置) 的变化作出合理的反应, 比如当人的移动速度变快 (或慢) 时机器人也应当适当地加快 (或减慢) 速度, 从而实现友好的交互环境.

为了避免时间同步问题, Ewerton 等[114] 假设在人机交互中人类的运动时长和训练样本中的时长是一样的. 然而正如文献 [20] 指出, 该假设在实际中是难以成立的. 因此 Maeda 等[20] 提出在 ProMP 中加入时间同步的方法, 其将人的运动轨迹和机器人的轨迹合并成更高维度的轨迹, 然后用式 (23) ~ (24) 获得合成轨迹对应的参数 $\boldsymbol{w}$ 的高斯分布. 该分布可以看作是人类运动轨迹参数 $\boldsymbol{w}_h$ 和机器人运动轨迹参数 $\boldsymbol{w}_r$ 的联合概率分布. 在人机交互时, 将人的轨迹实时的当作观测值并利用式 (25) ~ (26) 可对 $\boldsymbol{w}$ 进行更新[13], 这时 $\boldsymbol{w}$ 中的 $\boldsymbol{w}_r$ 即可用来生成机器人的轨迹. 最后, 文献 [20] 给出依据人的运动轨迹实时调整机器人运动时长的方法. 和文献 [20] 类似, Amor 等[115] 利用 DMP 分别对人类和机器人的轨迹进行学习, 并且给出人和机器人在时间上同步的方法. 其他需要时间同步的工作还包括应用 HMM 的方法[116].

如文献 [34] 中的分析, 上述方法在对人和机器人的轨迹进行建模时均采用时间作为轨迹的输入, 未能直接考虑人和机器人之间的协调关系, 故在预测机器人轨迹的同时需要附加的人机同步 (即在时间上) 的处理. 由于 KMP 可以学习带有多维输入的运动轨迹, Huang 等[6] 应用 KMP 直接根据人类的运动状态 (即输入) 对机器人的轨迹 (即输出) 进行预测, 后又将其推广到人机交互中机器人的姿态预测[34], 由于预测过程中不涉及时间, 避免了文献 [20, 114−116] 中的时间同步问题. 类似地, Silverio 等[117] 研究利用 GP 实现人机交互的问题, 然而该方法未考虑多维轨迹的协方差以及轨迹泛化问题. 另外, 基于动态系统的方法由于其直接对轨迹及其高阶微分进行学习[3, 118], 也能够避免人机交互中的时间同步问题.

4 讨论和展望

本节对模仿学习的一些未来发展趋势进行讨论和展望, 包括从轨迹规划的角度对模仿学习进行改进、结合任务分解和交互式反馈的模仿学习以及学习人类与环境交互过程中的因果关系.

4.1 结合轨迹优化或采样的模仿学习

在模仿学习之外, 轨迹规划 (Motion planning) 领域存在着大量的关于轨迹或路径规划的算法, 如第 3.6.1 节中提及的 CHOMP, 还有其他基于优化

[13] 该更新同时也需要机器人的观测轨迹, 然而该轨迹恰是需要预测的, 因此文献 [20] 在更新 $\boldsymbol{w}$ 时将机器人的观测值设成零向量, 同时将拟合机器人轨迹的基函数设成零矩阵.

的方法包括随机轨迹优化[119] (Stochastic trajectory optimization for motion planning, STOMP), 基于序列凸优化的 TrajOpt 算法[120] 和随机多模态轨迹优化[121] (Stochastic multimodal trajectory optimization, SMTO) 等, 以及基于采样的方法包括快速扩展随机树[122] (Rapidly-exploring random trees, RRT) 和概率路线图[123−124] (Probabilistic roadmap, PRM) 等. 模仿学习和这些方法的最大区别在于前者主要通过学习人类的示教轨迹达到模仿的效果, 而后者主要侧重快速的寻找满足任务或环境约束的可行轨迹. 目前相关的研究如第 3.6.2 节提及的文献 [106, 110] 可以看作是模仿学习和运动规划的结合, 然而两者在轨迹泛化以及复杂约束的情况下仍存在着很大的局限性. 因此如何将不同的轨迹规划算法和模仿学习进行有机的结合是未来研究的一个重要方向.

4.2 结合任务分解和交互式反馈的模仿学习

当面对复杂任务时, 人类可以直接地将其分解成一系列可行的子任务, 并且能够合理地分配各个子任务的难度以及子任务之间的协调. 对于机器人而言, 如何从 MP 库中选择恰当的 MP 以及对多个 MP 进行合理的串联是十分重要的. 如果采用强化学习的方法解决该问题, 则机器人将过于依赖与环境的交互且随着 MP 数量的增加其需要的训练次数也会显著地增加. 如果采用从示教轨迹中学习 MP 序列的方法 (如文献 [79]), 则只适用于和示教场景类似的情况, 无法泛化到更一般的未知问题. 因此研究人类对于不同任务或动作的分解和组合策略是模仿学习发展的另一个重要方向.

另外, 当 MP 库中的所有 MP 均无法或难以实现某个子任务时, 如 MP 库中的运动均为简单的点到点的运动而对于握手任务则需要周期运动, 如何引入人类的交互式反馈也是未来的一个重要研究方向. 目前已存在一些关于交互式学习的工作. 如文献 [6] 在 KMP 的框架下提出基于人机交互力的轨迹自适应的方法. 文献 [125] 研究通过人类的反馈对轨迹进行调整. 文献 [126] 利用 GP 预测的不确定性来判断是否需要人类提供新的示教样本. 然而文献 [6, 125−126] 均限于单独 MP 且应用对象仅为简单的任务 (如避障[6]、写字母[125] 和触碰动作[126]), 未涉及复杂任务的分解以及多 MP 的问题.

4.3 结合因果推理的模仿学习

对于人类技能的模仿学习除了学习轨迹本身还应考虑示教过程中蕴含的因果关系. 该关系可以认为是在抽象的层次对人类技能进行理解. 针对该问题, 可以采用因果推理 (Causal inference)[127] 提取观测变量间的因果关系和因果强度. 相关的研究如文献 [128], 其首先分析人类在对物体进行操作时意图之间的因果关系, 后将提取出的因果关系应用于新任务的泛化. 另外, 如文献 [129] 中的讨论, 在模仿学习中常常存在一些与人类行为决策无关的状态, 如果将这些状态应用于模仿学习将不利于技能的泛化, 而通过引入干涉 (Intervention) 的方法提取状态和行为之间正确的因果关系能够提高模仿学习的性能. 因此, 结合因果推理将是模仿学习研究的又一个重要趋势.

5 结论

本文介绍了模仿学习中的基本问题和主要方法, 并对其中各种方法的优点和局限性进行了讨论和比较. 在这些方法的基础上, 本文讨论了模仿学习中存在的若干关键问题. 另外, 本文探讨了未来可能的发展方向. 需要强调的是, 在实际机器人系统中模仿学习常和其他的算法紧密相连, 如文中提及的应用强化学习对运动基元进行优化、泛函梯度或随机采样和模仿学习相结合实现避障以及基于轨迹概率分布设计控制器等, 因此文中并未做严格区分.

References

1 Schaal S. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 1999, **3**(6): 233−242

2 Ijspeert A J, Nakanishi J, Hoffmann H, Pastor P, Schaal S. Dynamical movement primitives: Learning attractor models for motor behaviors. *Neural Computation*, 2013, **25**(2): 328−373

3 Khansari-Zadeh S M, Billard A. Learning stable nonlinear dynamical systems with Gaussian mixture models. *IEEE Transactions on Robotics*, 2011, **27**(5): 943−957

4 Paraschos A, Daniel C, Peters J, Neumann G. Probabilistic movement primitives. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. Nevada, USA: NIPS, 2013. 2616−2624

5 Calinon S, Bruno D, Caldwell D G. A task-parameterized probabilistic model with minimal intervention control. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 3339−3344

6 Huang Y L, Rozo L, Silverio J, Caldwell D G. Kernelized movement primitives. *The International Journal of Robotics Research*, 2019, **38**(7): 833−852

7 Muhlig M, Gienger M, Hellbach S, Steil J J, Goerick C. Task-level imitation learning using variance-based movement optimization. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Kobe, Japan: IEEE, 2009. 1177−1184

8 Huang Y L, Buchler D, Koc O, Scholkopf B, Peters J. Jointly learning trajectory generation and hitting point prediction in robot table tennis. In: Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots. Cancun, Mexico: IEEE, 2016. 650−655

9 Huang Y L, Silverio J, Rozo L, Caldwell D G. Hybrid probabilistic trajectory optimization using null-space exploration. In: Proceedings of the 2018 IEEE International Conference on Robotics and Automation. Brisbane, Australia: IEEE, 2018. 7226−7232

10 Stulp F, Theodorou E, Buchli J, Schaal S. Learning to grasp under uncertainty. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE, 2011. 5703−5708

11 Mylonas G P, Giataganas P, Chaudery M, Vitiello V, Darzi A, Yang G Z. Autonomous eFAST ultrasound scanning by a robotic manipulator using learning from demonstrations. In: Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013. 3251−3256

12 Reiley C E, Plaku E, Hager G D. Motion generation of robotic surgical tasks: Learning from expert demonstrations. In: Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. Buenos Aires, Argentina: IEEE, 2010. 967−970

13 Colome A, Torras C. Dimensionality reduction in learning Gaussian mixture models of movement primitives for contextualized action selection and adaptation. *IEEE Robotics and Automation Letters*, 2018, **3**(4): 3922−3929

14 Canal G, Pignat E, Alenya G, Calinon S, Torras C. Joining high-level symbolic planning with low-level motion primitives in adaptive HRI: Application to dressing assistance. In: Proceedings of the 2018 IEEE International Conference on Robotics and Automation. Brisbane, Australia: IEEE, 2018. 3273−3278

15 Joshi R P, Koganti N, Shibata T. A framework for robotic clothing assistance by imitation learning. *Advanced Robotics*, 2019, **33**(22): 1156−1174

16 Motokura K, Takahashi M, Ewerton M, Peters J. Plucking motions for tea harvesting robots using probabilistic movement primitives. *IEEE Robotics and Automation Letters*, 2020, **5**(2): 3275−3282

17 Ding J T, Xiao X H, Tsagarakis N, Huang Y L. Robust gait synthesis combining constrained optimization and imitation learning. In: Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE, 2020. 3473−3480

18 Zou C B, Huang R, Cheng H, Qiu J. Learning gait models with varying walking speeds. *IEEE Robotics and Automation Letters*, 2020, **6**(1): 183−190

19 Huang R, Cheng H, Guo H L, Chen Q M, Lin X C. Hierarchical interactive learning for a human-powered augmentation lower exoskeleton. In: Proceedings of the 2016 IEEE International Conference on Robotics and Automation. Stockholm, Sweden: IEEE, 2016. 257−263

20 Maeda G, Ewerton M, Neumann G, Lioutikov R, Peters J. Phase estimation for fast action recognition and trajectory generation in human-robot collaboration. *The International Journal of Robotics Research*, 2017, **36**(13−14): 1579−1594

21 Silverio J, Huang Y L, Abu-Dakka F J, Rozo L, Caldwell D G. Uncertainty-aware imitation learning using kernelized movement primitives. In: Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE, 2019. 90−97

22 Pomerleau D A. ALVINN: An autonomous land vehicle in a neural network. In: Proceedings of the 1st International Conference on Neural Information Processing Systems. Denver, USA: NIPS, 1989. 305−313

23 Ross S, Gordon G J, Bagnell D. A reduction of imitation learning and structured prediction to no-regret online learning. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA: JMLR.org, 2011. 627−635

24 Abbeel P, Ng A Y. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the 21st International Conference on Machine Learning. Banff, Canada: 2004. 1−8

25 Ho J, Ermon S. Generative adversarial imitation learning. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: NIPS, 2016. 4572−4580

26 Liu Nai-Jun, Lu Tao, Cai Ying-Hao, Wang Shuo. A review of robot manipulation skills learning methods. *Acta Automatica Sinica*, 2019, **45**(3): 458−470

27 Qin Fang-Bo, Xu De. Review of robot manipulation skill models. *Acta Automatica Sinica*, 2019, **45**(8): 1401−1418

28 Billard A, Epars Y, Cheng G, Schaal S. Discovering imitation strategies through categorization of multi-dimensional data. In: Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE, 2003. 2398−2403

29 Calinon S, Guenter F, Billard A. On learning, representing, and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2007, **37**(2): 286−298

30 Huang Y L, Abu-Dakka F J, Silverio J, Caldwell D G. Generalized orientation learning in robot task space. In: Proceedings of the 2019 International Conference on Robotics and Automation. Montreal, Canada: IEEE, 2019. 2531−2537

31 Matsubara T, Hyon S H, Morimoto J. Learning stylistic dynamic movement primitives from multiple demonstrations. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China: IEEE, 2010. 1277−1283

32 Giusti A, Zeestraten M J A, Icer E, Pereira A, Caldwell D G, Calinon S, et al. Flexible automation driven by demonstration: Leveraging strategies that simplify robotics. *IEEE Robotics & Automation Magazine*, 2018, **25**(2): 18−27

33 Huang Y L, Scholkopf B, Peters J. Learning optimal striking points for a ping-pong playing robot. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 2015. 4587−4592

34 Huang Y L, Abu-Dakka F J, Silverio J, Caldwell D G. Toward orientation learning and adaptation in Cartesian space. *IEEE Transactions on Robotics*, 2021, **37**(1): 82−98

35 Bishop C M. Pattern Recognition and Machine Learning. Heidelberg: Springer, 2006.

36 Cohn D A, Ghahramani Z, Jordan M I. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 1996, **4**: 129−145

37 Calinon S. A tutorial on task-parameterized movement learning and retrieval. *Intelligent Service Robotics*, 2016, **9**(1): 1−29

38 Guenter F, Hersch M, Calinon S, Billard A. Reinforcement learning for imitating constrained reaching movements. *Advanced Robotics*, 2007, **21**(13): 1521−1544

39 Peters J, Vijayakumar S, Schaal S. Natural actor-critic. In: Proceedings of the 16th European Conference on Machine Learning. Porto, Portugal: Springer, 2005. 280−291

40 Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, **77**(2): 257−286

41 Yu S Z. Hidden semi-Markov models. *Artificial Intelligence*, 2010, **174**(2): 215−243

42 Calinon S, D'halluin F, Sauser E L, Caldwell D G, Billard A G. Learning and reproduction of gestures by imitation. *IEEE Robotics & Automation Magazine*, 2010, **17**(2): 44−54

43 Osa T, Pajarinen J, Neumann G, Bagnell J A, Abbeel P, Peters J. An algorithmic perspective on imitation learning. *Foundations and Trends in Robotics*, 2018, **7**(1−2): 1−79

44 Zeestraten M J A, Calinon S, Caldwell D G. Variable duration movement encoding with minimal intervention control. In: Proceedings of the 2016 IEEE International Conference on Robotics and Automation. Stockholm, Sweden: IEEE, 2016. 497−503

45 Rasmussen C E, Williams C K I. Gaussian Processes for Machine Learning. Cambridge: MIT Press, 2006.

46 Hofmann T, Scholkopf B, Smola A J. Kernel methods in machine learning. *The Annals of Statistics*, 2008, **36**(3): 1171−1220

47 Alvarez M A, Rosasco L, Lawrence N D. Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, 2012, **4**(3): 195−266

48 Solak E, Murray-Smith R, Leithead W E, Leith D J, Rasmussen C E. Derivative observations in Gaussian process models of dynamic systems. In: Proceedings of the 15th International Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2002. 1057−1064

49 Atkeson C G, Moore A W, Schaal S. Locally weighted learning. *Artificial Intelligence Review*, 1997, **11**(1−5): 11−73

50 Kober J, Mulling K, Kromer O, Lampert C H, Scholkopf B, Peters J. Movement templates for learning of hitting and batting. In: Proceedings of the 2010 IEEE International Conference on Robotics and Automation. Anchorage, USA: IEEE, 2010. 853−858

51 Fanger Y, Umlauft J, Hirche S. Gaussian processes for dynamic movement primitives with application in knowledge-based cooperation. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon, Korea : IEEE, 2016. 3913−3919

52 Calinon S, Li Z B, Alizadeh T, Tsagarakis N G, Caldwell D G. Statistical dynamical systems for skills acquisition in humanoids. In: Proceedings of the 12th IEEE-RAS International Conference on Humanoid Robots. Osaka, Japan: IEEE, 2012. 323−329

53 Stulp F, Sigaud O. Robot skill learning: From reinforcement learning to evolution strategies. *Paladyn, Journal of Behavioral Robotics*, 2013, **4**(1): 49−61

54 Kober J, Oztop E, Peters J. Reinforcement learning to adjust robot movements to new situations. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain: IJCAI/AAAI, 2011. 2650−2655

55 Zhao T, Deng M D, Li Z J, Hu Y B. 2018. Cooperative manipulation for a mobile dual-arm robot using sequences of dynamic movement primitives. *IEEE Transactions on Cognitive and Developmental Systems*, 2020, **12**(1): 18−29

56 Li Z J, Zhao T, Chen F, Hu Y B, Su C Y, Fukuda T. Reinforcement learning of manipulation and grasping using dynamical movement primitives for a humanoidlike mobile manipulator. *IEEE/ASME Transactions on Mechatronics*, 2018, **23**(1): 121−131

57 Paraschos A, Rueckert E, Peters J, Neumann G. Model-free probabilistic movement primitives for physical interaction. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 2015. 2860−2866

58 Havoutis I, Calinon S. Supervisory teleoperation with online learning and optimal control. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017. 1534−1540

59 Hershey J R, Olsen P A. Approximating the Kullback Leibler divergence between Gaussian mixture models. In: Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing. Honolulu, USA: IEEE, 2007. IV-317−IV-320

60 Goldberg P W, Williams C K I, Bishop C M. Regression with input-dependent noise: A Gaussian process treatment. In: Proceedings of the 10th International Conference on Neural Information Processing Systems. Denver, USA: NIPS, 1998. 493−499

61 Kersting K, Plagemann C, Pfaff P, Burgard W. Most likely heteroscedastic Gaussian process regression. In: Proceedings of the 24th International Conference on Machine Learning. Corvalis, USA: ACM, 2007. 393−400

62 Pastor P, Hoffmann H, Asfour T, Schaal S. Learning and generalization of motor skills by learning from demonstration. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Kobe, Japan: IEEE, 2009. 763−768

63 Silverio J, Rozo L, Calinon S, Caldwell D G. Learning bimanual end-effector poses from demonstrations using task-parameterized dynamical systems. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 2015. 464−470

64 Ude A, Nemec B, Petric T, Morimoto J. Orientation in cartesian space dynamic movement primitives. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 2997−3004

65 Abu-Dakka F J, Nemec B, Jorgensen J A, Savarimuthu T R, Kruger N, Ude A. Adaptation of manipulation skills in physical contact with the environment to reference force profiles. *Autonomous Robots*, 2015, **39**(2): 199−217

66 Ravichandar H, Dani A. Learning position and orientation dynamics from demonstrations via contraction analysis. *Autonomous Robots*, 2019, **43**(4): 897−912

67 Zeestraten M J A, Havoutis I, Silverio J, Calinon S, Caldwell D G. An approach for imitation learning on Riemannian manifolds. *IEEE Robotics and Automation Letters*, 2017, **2**(3): 1240−1247

68 Saveriano M, Franzel F, Lee D. Merging position and orientation motion primitives. In: Proceedings of the 2019 International Conference on Robotics and Automation. Montreal, Canada: IEEE, 2019. 7041−7047

69 Abu-Dakka F J, Kyrki V. Geometry-aware dynamic movement primitives. In: Proceedings of the 2020 IEEE International Conference on Robotics and Automation. Paris, France: IEEE, 2020. 4421−4426

70 Abu-Dakka F J, Huang Y L, Silverio J, Kyrki V. A probabilistic framework for learning geometry-based robot manipulation skills. *Robotics and Autonomous Systems*, 2021, **141**: 103761

71 Calinon S. Gaussians on Riemannian manifolds: Applications for robot learning and adaptive control. *IEEE Robotics & Automation Magazine*, 2020, **27**(2): 33−45

72 Kronander K, Billard A. Learning compliant manipulation through kinesthetic and tactile human-robot interaction. *IEEE Transactions on Haptics*, 2014, **7**(3): 367−380

73 Wu Y Q, Zhao F, Tao T, Ajoudani A. A framework for autonomous impedance regulation of robots based on imitation learning and optimal control. *IEEE Robotics and Automation Letters*, 2021, **6**(1): 127−134

74 Forte D, Gams A, Morimoto J, Ude A. On-line motion synthesis and adaptation using a trajectory database. *Robotics and Autonomous Systems*, 2012, **60**(10): 1327−1339

75 Kramberger A, Gams A, Nemec B, Chrysostomou D, Madsen O, Ude A. Generalization of orientation trajectories and force-torque profiles for robotic assembly. *Robotics and Autonomous Systems*, 2017, **98**: 333−346

76 Stulp F, Raiola G, Hoarau A, Ivaldi S, Sigaud O. Learning compact parameterized skills with a single regression. In: Proceedings of the 13th IEEE-RAS International Conference on Humanoid Robots. Atlanta, USA: IEEE, 2013. 417−422

77 Huang Y L, Silverio J, Rozo L, Caldwell D G. Generalized task-parameterized skill learning. In: Proceedings of the 2018 IEEE International Conference on Robotics and Automation. Brisbane, Australia: IEEE, 2018. 5667−5474

78 Kulic D, Ott C, Lee D, Ishikawa J, Nakamura Y. Incremental

learning of full body motion primitives and their sequencing through human motion observation. *The International Journal of Robotics Research*, 2012, **31**(3): 330−345

79 Manschitz S, Gienger M, Kober J, Peters J. Learning sequential force interaction skills. *Robotics*, 2020, **9**(2): 45

80 Kober J, Gienger M, Steil J J. Learning movement primitives for force interaction tasks. In: Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Seattle, USA: IEEE, 2015. 3192−3199

81 Medina J R, Billard A. Learning stable task sequences from demonstration with linear parameter varying systems and hidden Markov models. In: Proceedings of the 1st Annual Conference on Robot Learning. Mountain View, USA: PMLR, 2017. 175−184

82 Meier F, Theodorou E, Stulp F, Schaal S. Movement segmentation using a primitive library. In: Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA: IEEE, 2011. 3407−3412

83 Lee S H, Suh I H, Calinon S, Johansson R. Autonomous framework for segmenting robot trajectories of manipulation task. *Autonomous Robots*, 2015, **38**(2): 107−141

84 Stulp F, Schaal S. Hierarchical reinforcement learning with movement primitives. In: Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots. Bled, Slovenia: IEEE, 2011. 231−238

85 Daniel C, Neumann G, Kroemer O, Peters J. Learning sequential motor tasks. In: Proceedings of the 2013 IEEE International Conference on Robotics and Automation. Karlsruhe, Germany: IEEE, 2013. 2626−2632

86 Duan A Q, Camoriano R, Ferigo D, Huang Y L, Calandriello D, Rosasco L, et al. Learning to sequence multiple tasks with competing constraints. In: Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE, 2019. 2672−2678

87 Silverio J, Huang Y L, Rozo L, Calinon S, Caldwell D G. Probabilistic learning of torque controllers from kinematic and force constraints. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain: IEEE, 2018. 1−8

88 Schneider M, Ertel W. Robot learning by demonstration with local Gaussian process regression. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China: IEEE, 2010. 255-260

89 Umlauft J, Fanger Y, Hirche S. Bayesian uncertainty modeling for programming by demonstration. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017. 6428−6434

90 Wilson A G, Ghahramani Z. Generalised Wishart processes. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. Barcelona, Spain: AUAI Press, 2011. 1−9

91 Medina J R, Lee D, Hirche S. Risk-sensitive optimal feedback control for haptic assistance. In: Proceedings of the 2012 IEEE International Conference on Robotics and Automation. Saint Paul, USA: IEEE, 2012. 1025−1031

92 Huang Y L, Silverio J, Caldwell D G. Towards minimal intervention control with competing constraints. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain: IEEE, 2018. 733−738

93 Calinon S, Billard A. A probabilistic programming by demonstration framework handling constraints in joint space and task space. In: Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice, France: IEEE, 2008. 367−372

94 Calinon S, Billard A. Statistical learning by imitation of competing constraints in joint space and task space. *Advanced Robotics*, 2009, **23**(15): 2059−2076

95 Paraschos A, Lioutikov R, Peters J, Neumann G. Probabilistic prioritization of movement primitives. *IEEE Robotics and Automation Letters*, 2017, **2**(4): 2294−2301

96 Fajen B R, Warren W H. Behavioral dynamics of steering, obstable avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, 2003, **29**(2): 343−362

97 Hoffmann H, Pastor P, Park D H, Schaal S. Biologically-inspired dynamical systems for movement generation: Automatic real-time goal adaptation and obstacle avoidance. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Kobe, Japan: IEEE, 2009. 2587−2592

98 Duan A Q, Camoriano R, Ferigo D, Huang Y L, Calandriello D, Rosasco L, et al. Learning to avoid obstacles with minimal intervention control. *Frontiers in Robotics and AI*, 2020, **7**: 60

99 Park D H, Hoffmann H, Pastor P, Schaal S. Movement reproduction and obstacle avoidance with dynamic movement primitives and potential fields. In: Proceedings of the 8th IEEE-RAS International Conference on Humanoid Robots. Daejeon, Korea: IEEE, 2008. 91−98

100 Maciejewski A A, Klein C A. Obstacle avoidance for kinematically redundant manipulators in dynamically varying environments. *The International Journal of Robotics Research*, 1985, **4**(3): 109−117

101 Shyam R B, Lightbody P, Das G, Liu P C, Gomez-Gonzalez S, Neumann G. Improving local trajectory optimisation using probabilistic movement primitives. In: Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE, 2019. 2666−2671

102 Zucker M, Ratliff N, Dragan A D, Pivtoraiko M, Klingensmith M, Dellin C M, et al. CHOMP: Covariant hamiltonian optimization for motion planning. *The International Journal of Robotics Research*, 2013, **32**(9−10): 1164−1193

103 Huang Y L, Caldwell D G. A linearly constrained nonparametric framework for imitation learning. In: Proceedings of the 2020 IEEE International Conference on Robotics and Automation. Paris, France: IEEE, 2020. 4400−4406

104 Saveriano M, Lee D. Learning barrier functions for constrained motion planning with dynamical systems. In: Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE, 2019. 112−119

105 Huang Y L. EKMP: Generalized imitation learning with adaptation, nonlinear hard constraints and obstacle avoidance. arXiv: 2103.00452, 2021.

106 Osa T, Esfahani A M G, Stolkin R, Lioutikov R, Peters J, Neumann G. Guiding trajectory optimization by demonstrated distributions. *IEEE Robotics and Automation Letters*, 2017, **2**(2): 819−826

107 Marinho Z, Boots B, Dragan A, Byravan A, Srinivasa S, Gordon G J. Functional gradient motion planning in reproducing kernel Hilbert spaces. In: Proceedings of the Robotics: Science and Systems XII. Ann Arbor, USA, 2016. 1−9

108 Rana M A, Mukadam M, Ahmadzadeh S R, Chernova S, Boots B. Towards robust skill generalization: Unifying learning from demonstration and motion planning. In: Proceedings of the 1st Annual Conference on Robot Learning. Mountain View, USA: PMLR, 2017. 109−118

109 Koert D, Maeda G, Lioutikov R, Neumann G, Peters J. Demonstration based trajectory optimization for generalizable robot motions. In: Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots. Cancun, Mexico: IEEE, 2016. 515−522

110 Ye G, Alterovitz R. Demonstration-guided motion planning. Robotics Research. Cham: Springer, 2017. 291−307

111 Englert P, Toussaint M. Learning manipulation skills from a single demonstration. *The International Journal of Robotics Research*, 2018, **37**(1): 137−154

112 Doerr A, Ratliff N D, Bohg J, Toussaint M, Schaal S. Direct

loss minimization inverse optimal control. In: Proceedings of the Robotics: Science and Systems. Rome, Italy, 2015. 1−9

113 Hansen N. The CMA evolution strategy: A comparing review. *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms*. Berlin, Heidelberg: Springer, 2006, 75−102

114 Ewerton M, Neumann G, Lioutikov R, Amor H B, Peters J, Maeda G. Learning multiple collaborative tasks with a mixture of interaction primitives. In: Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Seattle, USA: IEEE, 2015. 1535−1542

115 Amor H B, Neumann G, Kamthe S, Kroemer O, Peters J. Interaction primitives for human-robot cooperation tasks. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 2831−2837

116 Vogt D, Stepputtis S, Grehl S, Jung B, Amor H B. A system for learning continuous human-robot interactions from human-human demonstrations. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017. 2882−2889

117 Silverio J, Huang Y L, Rozo L, Caldwell D G. An uncertainty-aware minimal intervention control strategy learned from demonstrations. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain: IEEE, 2018. 6065−6071

118 Khoramshahi M, Billard A. A dynamical system approach to task-adaptation in physical human-robot interaction. *Autonomous Robots*, 2019, **43**(4): 927−946

119 Kalakrishnan M, Chitta S, Theodorou E, Pastor P, Schaal S. STOMP: Stochastic trajectory optimization for motion planning. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE, 2011. 4569−4574

120 Schulman J, Duan Y, Ho J, Lee A, Awwal I, Bradlow H, et al. Motion planning with sequential convex optimization and convex collision checking. *The International Journal of Robotics Research*, 2014, **33**(9): 1251−1270

121 Osa T. Multimodal trajectory optimization for motion planning. *The International Journal of Robotics Research*, 2020, **39**(8): 983−1001

122 LaValle S M, Kuffner Jr J J. Randomized kinodynamic planning. *The International Journal of Robotics Research*, 2001, **20**(5): 378−400

123 Kavraki L E, Svestka P, Latombe J C, Overmars M H. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*, 1996, **12**(4): 566−580

124 Hsu D, Latombe J C, Kurniawati H. On the probabilistic foundations of probabilistic roadmap planning. *The International Journal of Robotics Research*, 2006, **25**(7): 627−643

125 Celemin C, Maeda G, Ruiz-del-Solar J, Peters J, Kober J. Reinforcement learning of motor skills using policy search and human corrective advice. *The International Journal of Robotics Research*, 2019, **38**(14): 1560−1580

126 Maeda G, Ewerton M, Osa T, Busch B, Peters J. Active incremental learning of robot movement primitives. In: Proceedings of the 1st Annual Conference on Robot Learning. Mountain View, USA: PMLR, 2017. 37−46

127 Pearl J. *Causality*. Cambridge: Cambridge University Press, 2009.

128 Katz G, Huang D W, Hauge T, Gentili R, Reggia J. A novel parsimonious cause-effect reasoning algorithm for robot imitation and plan recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 2018, **10**(2): 177−193

129 Haan P, Jayaraman D, Levine S. Causal confusion in imitation learning. In: Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada: NeurIPS, 2019. 11693−11704

黄艳龙 英国利兹大学计算机系助理教授. 主要研究方向为模仿学习, 强化学习和运动规划. 本文通信作者. E-mail: y.l.huang@leeds.ac.uk
(**HUANG Yan-Long** University academic fellow at the School of Computing, University of Leeds, Leeds, UK. His interest covers imitation learning, reinforcement learning and motion planning. Corresponding author of this paper.)

徐 德 中国科学院自动化研究所研究员. 1985 年、1990 年获得山东工业大学学士、硕士学位. 2001 年获得浙江大学博士学位. 主要研究方向为机器人视觉测量, 视觉控制, 智能控制, 视觉定位, 显微视觉, 微装配. E-mail: de.xu@ia.ac.cn
(**XU De** Professor at the Institute of Automation, Chinese Academy of Sciences. He received his bachelor and master degrees from Shandong University of Technology in 1985 and 1990, respectively. He received his Ph. D. degree from Zhejiang University in 2001. His research interest covers robotics and automation, such as visual measurement, visual control, intelligent control, visual positioning, microscopic vision, and microassembly.)

谭 民 中国科学院自动化研究所复杂系统管理与控制国家重点实验室研究员. 主要研究方向为机器人系统和智能控制系统. E-mail: min.tan@ia.ac.cn
(**TAN Min** Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interest covers robotics and intelligent control systems.)