

面向对手建模的意图识别方法综述

高巍^{1,2}, 罗俊仁¹, 袁唯淋¹, 张万鹏¹

(1. 国防科技大学智能科学学院, 湖南 长沙 410073;

2. 96843 部队, 甘肃 兰州 730102)

摘要: 首先介绍了对手建模的几种不同的类型, 引出行为建模中的意图识别问题; 随后针对意图识别的过程、分类、主要研究方法、研究展望以及实际应用进行了归纳分析, 总结并讨论了相关领域取得的最新研究成果; 最后指出意图识别目前存在的不足以及未来的发展方向。

关键词: 对手建模; 意图识别; 目标识别; 计划识别; 目标识别设计; 计划识别设计

中图分类号: TP18

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2021052

Survey of intention recognition for opponent modeling

GAO Wei^{1,2}, LUO Junren¹, YUAN Weilin¹, ZHANG Wanpeng¹

1. College of Intelligence Science and Technology, National University of Defense and Technology, Changsha 410073, China

2. The army of 96843, Lanzhou 730102, China

Abstract: Several different methods of opponent modeling were introduced, leading to the problem of intention recognition in behavior modeling. Then, the process, classification, main methods, research prospects and practical applications of intention recognition were analyzed inductively, the latest research in related fields were summarized. Finally, some shortcomings of the current intention recognition and design methods were pointed out and some new insights for the future research were presented.

Keywords: opponent modeling, intention recognition, goal recognition, plan recognition, goal recognition design, plan recognition design

1 引言

基于行为观测序列推断智能体的目标是近几年来人工智能和心理学研究者都感兴趣的问题。

众多行为建模方法中, 人工智能领域关于行为识别与设计得到了广泛关注。计划行动与意图识别(PAIR, plan, activity, and intent recognition)^[1]作为一种基于经典识别范式得到了广泛研究。

收稿日期: 2020-07-10; 修回日期: 2020-09-25

通信作者: 张万鹏, wpzhang@nudt.edu.cn

基金项目: 国家自然科学基金(61702528, 61806212, U1734208)

Foundation Item: The National Natural Science Foundation of China (61702528, 61806212, U1734208)

论文引用格式: 高巍, 罗俊仁, 袁唯淋, 等. 面向对手建模的意图识别方法综述[J]. 网络与信息安全学报, 2021, 7(4): 86-100.

GAO W, LUO J R, YUAN W L, et al. Survey of intention recognition for opponent modeling[J]. Chinese Journal of Network and Information Security, 2021, 7(4): 86-100.

目标识别被广泛应用于人机交互^[2]、智能体建模^[3]、目标推理^[4]、重要基础设施保护^[5]等。一些新的研究包括将意图识别应用于多智能体行为识别^[6]、决策过程中控制意图的分享与隐藏^[7]，之间的关系可分为：不可知识识别（agnostic/keyhole recognition），其中执行者对观察者的识别过程不可知；对抗识别，执行者在给定约束下试图欺骗识别者；有意识识别，执行者有意通过隐式通信的方式将自己的行为告知识别者。这些与可解释性规划^[8]、隐私保护规划密切相关^[7,9]，其中任务执行者可根据相关情景主动披露或隐秘混淆其真实目标。

2 对手建模

对手建模是在对抗环境下，考虑如何对除自己以外其他参与者进行行为建模，这是一种典型的行为预测技术^[10-11]。当前对手建模（OM，opponent modeling）^[12]的主要方法有以下几种（如图1所示）。

（1）策略重构（policy reconstruction）^[13]，假设模型有固定的结构，可根据被观察智能体行为，预测行为概率；可以学习满足模型结构的任意模型，交互过程中模型持续更新，但这种方式比较复杂，既耗时又耗空间，可细分为条件行为概率模型^[14]、案例推理模型^[15]、紧致模型表示^[16]和效用重构模型^[17]。

（2）类型推理（type reasoning）^[18-20]，假设智能体有多种类型，可根据被观察智能体行为分析模型的相对可能性，预测行为概率，其中“类型”可以是“黑盒”，如果真实类型存在可快速匹配，但类型空间出错时，行为预测可能出错，信念空间不足以分辨类型的真假。

（3）行为分类（action classification）^[21]，根据不同信息源信息，选择模型结构，运用机器学习的方法拟合模型参数，预测智能体的类别，可预测模型的多种属性，许多机器学习方法可能使用，但需要大量数据，模型需要提前计算，实时交互更新很难。

（4）行为识别（action recognition）^[1]，算法使用层次化规划库或域理论（domain theory），预测智能体的目标和可能行为，丰富的规划库有助

于识别复杂行为，但规划库可能不完备，许多识别方法没有考虑观察者的存在。

（5）递归推理（recursive reasoning）^[22-23]，递归持续推理其他智能体的可能模型，预测其可能行为，可用于分析其他智能体的高层次信念，但递归层次难以分析，理性假设太强。

（6）图模型（graph model）^[24-25]，使用各种图模型表示智能体决策过程与偏好，预测智能体的可能行为，图形化表示有助于计算效率的提高，但对于序列决策过程扩展性不强。

（7）群组建模（group modeling）^[26-27]，通过分析一群智能体的行动、目标和规划，预测多智能体的联合属性，可抓住一群智能体行为的关联关系，但智能体之间的不相关性，使群组建模很困难。

（8）集群建模（swarm modeling）^[28]，通过分析生物群体的行为特征对集群行为模型进行建模。

除以上介绍以外，隐式建模（implicit modeling）^[29]、假设检验^[30]和安全最佳反应^[31]也常用于智能体行为建模。



图1 对手建模分类
Figure 1 Classifications of opponent modeling

当前这类行为建模方法广泛应用于游戏AI中^[32-33]。智能体行为建模面临的一些问题主要有建模方法的组合、部分可观情况下的策略重构、安全有效的模型探索、决策因素的有效发觉、算法的高效实现、时变行为的建模、持续行为的建模、开放域多智能体建模、自主模型补全及修正等。

3 意图识别

3.1 意图识别过程

当前研究智能体意图的两大模型主要是 BDI

(信念、愿望和意图)模型^[29]和 COI (能力、机会和意图)模型^[5]。意图识别问题的三大研究要素为环境(或问题域)、识别者(或识别系统)以及被识别者。传统的行为识别方法通常将被识别者和识别者分别设定为行为执行者与观察者。

如图 2 所示,意图识别可以看作一个以观测信息为输入,在领域知识支撑下围绕假说空间的推理与计算,以符合观测的假说(包括目标与计划)及其概率分布为输出的过程。

3.2 意图识别分类

根据环境(或问题域)的类型,可以分为确定性识别和随机性识别。根据执行者和观察者之间的竞合关系,可以分为:不可知识别,其中执行者不知道观察者的识别过程;有意识识别,执行者有意通过隐式通信的方式将自己的行为告知识别者;对抗识别,执行者在给定约束下试图欺骗识别者^[1,5]。

意图识别作为“计划行动意图识别”的子问题,与计划识别、行动识别^[34]和目标识别紧密相关。“计划行动意图识别”有许多共同之处,为了充分发挥不同类型识别方法的作用,Freedman 等^[35]

提出“计划行动意图识别”的统一框架。根据研究问题的层次划分,当前的研究大多把行动识别看作底层识别,把计划识别看作包括连续行动和目标的识别,把目标识别仅看作识别行动目标。其实,当智能体的行动与具体特定目标关联时,意图识别即目标识别。当智能体处于确定性环境(或问题域)时,其选择的行动是依经典规划理论得出与特定目标关联的计划,包含为了达到目标而执行全部行动和行动间的关系,意图识别即计划识别。当智能体处于随机性环境(或问题域)时,其选择的行动是依序贯决策理论或博弈理论得出与特定目标相关联的行动,意图识别即目标识别。本文主要研究目标识别与计划识别。

3.3 意图识别方法

本文研究假定执行意图识别过程中,观察者充当识别者。行动识别主要识别底层行为,计划识别主要识别行为的中上层,目标识别主要识别行为的最终目标。

当前意图识别主要有以下三大范式。

(1) 基于规划理论的识别范式。这是一种典型的符号式确定性意图识别方法,其中可以用分

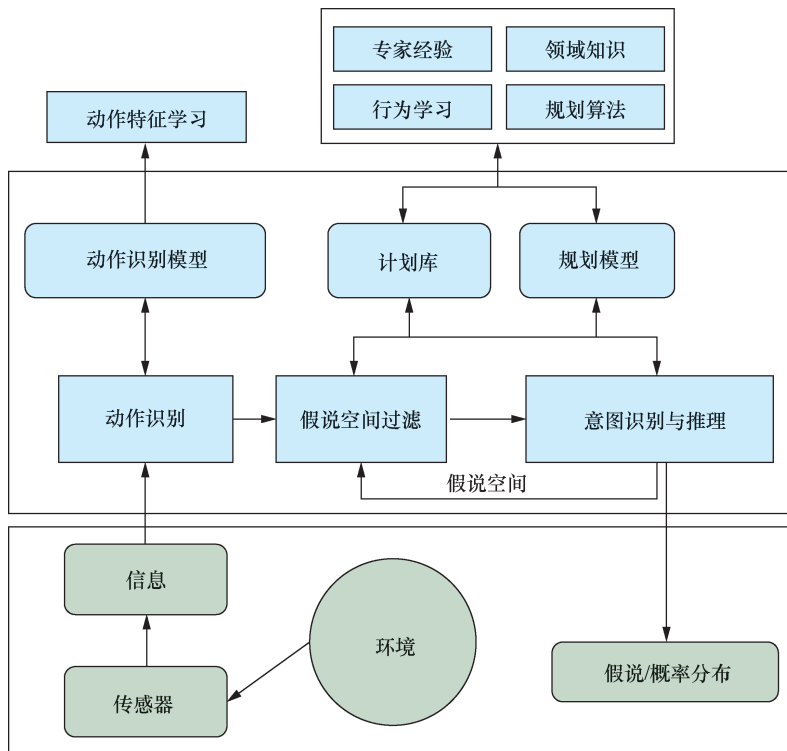


图 2 意图识别行为建模框架
Figure 2 Intention recognition behavior modeling framework

层任务网络 (HTN, hierarchical task network)、命题语法、与或树的方式来构建执行者的计划库或直接使用域理论来表征执行者的行为。基于规划的计划识别方法 (PRAP)^[36-37] 是一种利用逆向推理识别意图的方法, 其相应变体如基于规划的目标识别 (GRAP)^[38]、基于规划的模型识别 (MRAP)^[39] 均是基于域理论来进行行为识别的。这类方法大多依赖于计划库或域理论生成被观察到的智能体为实现其目标的计划假设空间^[1]。此类方法在计算上是有效的, 但需要丰富的领域知识, 并对观察到的智能体的偏好做出强有力的假设。

(2) 基于效用理论的识别范式^[5,38,40]。其中, 基于部分可观马尔可夫决策过程 (POMDP) 的识别方法^[38], 建立了不确定条件下的序贯决策模型, 为随机环境意图识别提供了方法; 而基于博弈理论的识别方法^[5,40], 使用一种生成式方法^[41], 为对抗博弈场景下意图识别与应对规划提供了新的模型与求解方法。

(3) 基于学习理论的识别范式。其中, 策略识别方法^[42-43]、逆强化学习^[44-46]和正则化强化学习^[47-48]方法均可用于学习行为模型、基于代价的深度学习^[49-52]方法, 可用于预测后续行动。

3.3.1 基于代价的目标识别

目标识别是识别目标驱动智能体行动策略的

一种主要方法, 这种方法主要着眼于智能体的目标域相关问题, 相较于计划识别中注重过程信息, 目标域问题将过程随机化, 重点关注目标的状态, 当前关于目标域攻防对抗问题的相关研究已经取得相当多的成果。

目标识别作为“计划行动意图识别”的子问题, 广泛应用于环境部分可观的决策问题, 其中观察者需要构造执行者的决策行动模型, 基于不完全观测序列识别执行者的行动目标。

目标识别通过观察智能体的行为来确定其意图的问题, 涉及对最终或“最高层次”目标感兴趣的配对问题, 而不是必须在实现计划或次级目标的过程中进行谈判的计划或次级目标^[53]。

如图3所示, 识别者对所感知的智能体行为进行分析, 确定智能体的目标。智能体以初始状态 (标记为“开始”) 进入场景, 并执行指向预先确定目标的一系列操作 (在这种情况下, 由图中标有“目标”的下部方框表示), 根据感知到的观察序列 (不一定完整或准确) 分析智能体的目标。此任务与计划识别紧密相关, 但又不同于计划识别, 后者旨在确定智能体要遵循的完整计划以实现其目标。

根据时效性要求可划分为离线目标识别和在线目标识别 (如表1所示)。在离线目标识别的观

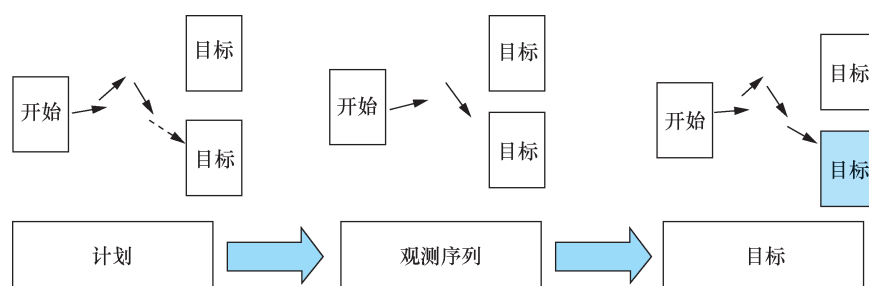


图3 目标识别过程

Figure 3 Target recognition process

表1 目标识别模型属性
Table 1 Target recognition model attributes

目标识别模型	原理	方法	连续	离散	在线	离线
基于代价的目标识别	根据状态与目标状态之间的代价差	R&G		✓		✓
		M&S	✓	✓		✓
		V&K	✓	✓		✓
基于计划度量的目标识别	根据计划执行过程进行分析方法	Landmark	✓		✓	✓

测序列中, 虽然本身可能会有很多噪声或者是不完整的, 但观测序列会提前显示出来。相反, 在线目标识别中, 观测值是逐步显示的, 而必须在每一次额外的观察之后做出假设, 且不知道哪一次观察是最后一次观察。当前, 离线目标识别研究比较多^[54], 近些年也出现了一些在线目标识别相关方法^[55-57]。

当前基于代价的目标识别模型主要有: R&G (Ramirez and Geffner) 模型^[36], 可用于与任务规划相关的经典离线目标识别方法; M&S (Masters and Sardina) 模型^[58], 可用于连续和离散空间与路径规划相关的离线目标识别方法; V&K (Vered and Kaminka) 模型^[55], 可用于连续和离散空间与运行规划相关的在线目标识别方法。

R&G 方法: Ramirez 和 Geffner 介绍了第一个基于计划的目标识别 (R&G) 方法。给定理论 D 和观测序列 O , R&G 引入了一种考虑观测序列 O 的修正理论。

给定 D 、 O 的修正理论 $D' = \langle S, A', G, I, C \rangle$,

其中每一个行动 $a \in A$, 在其效果列表中, 带有可能谓词的平行行动 A' 。例如, $p_b \rightarrow p_a$, a 是 O 的第一个观测; b 是紧挨着观测 O 中行动 a 的前一个行动。

给定一个计划 π , 当且仅当 π 满足 O 中的所有观测值直到 a 时, 表明 p_a 成立。R&G 方法使用 D , 使用现成的经典规划器来计算以下两个量。

1) 根据观测序列 O , 到达目标 g_i 的最小代价为 $C_i(O)$ 。

2) 忽略观测序列 O , 到达目标 g_i 的最小代价为 $C_i(-O)$ 。

$L(g_i | O)$ 是指直接以观察为基础的计划与达到 g_i 的最佳计划之间的邻近程度 (按代价计算), 定义为

$$L(g_i | O) = C_i(O) - C_i(-O)$$

随着区别的增大, 增加了给定观测序列 O 中目标 g_i 的可能性。给出一组观测序列 $O = \{o_1, \dots, o_i\}$, R&G 方法通过在识别阶段对给定 O 的每个目标调用一次规划器来计算 $C_i(O)$ 和 $C_i(-O)$ 这两个代价。

给定观察序列 O 的目标 g_i 的得分定义为

$$p(g_i | O) \cong \frac{1}{e^{\beta \cdot L(g_i | O)} + 1}$$

其中, $0 \leq \beta \leq 1$ 用于缓和偏离最优计划的观测影响。在计算了所有 $g_i \in G$ 的 $p(g_i | O)$ 后, 对它们进行归一化处理, 得到有效的概率分布。

M&S 方法: Masters 和 Sardina 建议用 C_i 取代 $C_i(-O)$, 在没有参考观测的情况下实现目标 g_i 的最低成本, 同时, 可以通过脱机调用规划器来计算, 而无须引用观测结果。由此得出的可能性公式如下。

$$L(g_i | O) = C_i(O) - C_i$$

事实证明, 使用 M&S 方法可以显著缩短计算时间, 同时具有与 R&G 方法相似的性能。

V&K 方法: 根据观测序列 O , 到达目标 g_i 的最小代价记为 $C_i(O)$, 忽略观测序列 O 到达目标 g_i 的最小代价记为 $c_i(s, g_i)$, 则目标概率为 $p(s, O, g_i) = \alpha \frac{c_i(s, g_i)}{c_i(s, O, g_i)}$, 其中 α 是归一化参数。

近些年, 关于温度参数的研究得到了广泛关注^[58], 主要是通过设计温度参数的取值进而优化识别模型。 $\beta = \frac{1}{\min(c_i(-O), c_i(O))}$ 与 R&G 模型

结合, 如针对非理性行为目标识别^[59], 取

$$RM(s, G, O) = \max_{g \in G} \frac{c_i(s, g_i)}{c_i(s, O, g_i)}$$

$\beta = RM^\gamma$ 作为自省模型的一部分与 M&S 模型结合。

3.3.2 基于计划度量的目标识别

在规划理论中, landmark 是指在每一个有效的计划中, 为了从最初的状态达到一个特定的目标, 必须在某一点上都是正确的事实^[60]。

Vered 和 Kaminka 进行了一系列关于连续域上的在线识别工作^[55-57], 考虑与路线规划以外的运动规划类型有关的目标识别, 如绘图分析。同时, 将目标识别描述为“目标镜像”(即人类对观察的移情反应)^[55], 即人类在观察到的情况下想象自己, 并假定被观察的人有兴趣发现最符合人类推理的“启发式”(即概率分布)。在关注在线目标识别的机制时, Vered 和 Kaminka 通过重复使用“路径前缀”(即到目前为止观察到的路径)的计算代价来节省时间, 而不是重复计算其全部

代价, 进而提出了两种新的机制, 以帮助减少计算代价^[56]。在在线目标识别中, 对观测数据进行增量处理, 在收到新的观测结果后, 该系统首先检查对手是否在接近“最有可能”的目标。如果是, 则跳过重新计算; 如果不是, 系统将检查自上一次观测以来对手的行为轨迹(即运动角度)。如果对手以大于某个给定阈值的角度偏离了一个目标, 则该目标将从候选集中剪枝, 进行未来概率计算。Vered 和 Kaminka 将目标镜像与 landmark 的使用结合起来^[57], 为离散域和连续域提供在线解决方案, 通过每个目标的“landmark 完成”与“总 landmark”的比率来生成概率分布。

3.3.3 计划识别

计划识别是许多领域中普遍存在的任务, 包括自然语言、多智能体系统和辅助认知^[61-63]。

计划识别是反向规划: 在规划中, 寻求实现目标的行动; 在计划识别中, 寻求可以解释观察到行动的目标。二者是相互联系的, 都需要采取一些假设来解释给定的数据: 用计划来解释目标, 用目标来解释部分观察到的计划。在 Ramírez 和 Geffner 工作中^[36], 通过利用经典规划算法的作用和通用性来缩小计划识别和规划之间的差距。为此, 他们从计划识别问题转移到计划库和涉及领域理论以及目标的可能集合 $\mathcal{G}$ 的计划识别问题。然后, 将计划识别任务表述为识别目标 $G \in \mathcal{G}$ 的问题, 以使 G 的某个最佳计划与观测值兼容。之所以关注 G 的最优计划, 是因为它们代表了追求目标 G 的完全理性的智能体可能选择的行为。

计划识别是根据智能体观察到的动作的不完整序列推断其计划的任务。给定一个包含所有可能目标集 $\mathcal{G}$ 的计划库的计划识别问题, 可以抽象地理解为在库中找到实现可能目标集中的任意目标 G 计划 π , 记为 $\pi \in \Pi_L(G)$, 其中, 计划 π 满足观察得到的结果。通过域理论以类似的方式定义计划识别问题, 只是在给定域 P 的情况下, 将 G 中最佳计划的集合 $\pi \in \Pi_P^*(G)$ 更改为 G 中最佳计划的集合 $\pi \in \Pi_L(G)$ 。使用 $P = \langle F, O, I \rangle$ 表示规划领域, 以便通过将规划领域与目标 G 相连接来获得规划问题 $P = \langle G \rangle$ 。

计划识别问题可以用一个三元组 $T = \langle P, O, \mathcal{G} \rangle$ 表示, 其中 $P = \langle F, O, A \rangle$ 是计划领

域, $\mathcal{G}$ 是所有可能目标 G 的集合, $\mathcal{G} \subseteq F$ 和 $O = \{o_1, o_2, \dots, o_m\}$ 是观测序列, 每个序列都在动作 A 中。

其中, 选取的行动序列满足由动作组成的观察序列。例如, 动作序列 $\pi = \{a, b, c, d, e, a\}$ 满足观察序列 $O_1 = \{b, d, a\}$ 和 $O_2 = \{a, c, a\}$, 但不是 $O_3 = \{b, d, c\}$ 。这可以借助将 O 中的观察索引映射为 A 中的动作索引的函数来形式化。

一个动作序列 $A = \{a_1, a_2, \dots, a_m\}$ 满足观测序列 $O = \{o_1, o_2, \dots, o_m\}$, 如果有一个单调函数 f 映射观测指数 $j = \{1, 2, \dots, m\}$ 成动作指数 $i = \{1, 2, \dots, n\}$, 那么 $a_f(j) = O_i$ 。

3.4 意图识别设计

3.4.1 目标识别设计

目标识别设计 (GRD, goal recognition design) 由 Keren、Gal 和 Karpas 在 2014 年提出^[60], 其目的是通过离线执行目标识别任务来降低在线目标识别任务的复杂性, 修改智能体操作的底层环境。目标识别设计的目的是找到最小的一组底层环境的修改, 迫使智能体尽早揭示其目标。这个问题本身与许多相同的目标识别应用程序相关, 因为通常可以很容易修改底层环境。

通常情况下, 目标识别设计问题有两个组成部分: 目标识别效果的度量; 可能对环境造成的设计变化的模型。

目前, 目标识别设计模型主要包括确定性的目标识别设计与随机性的目标识别设计, 表 2 总结了现有目标识别设计模型的概括、度量和可能的的设计。

(1) 确定性目标识别设计

目标识别设计主要分析如何更好地改造执行者所处的环境, 使执行者的目标更早暴露出来, 其中最坏情况显著性 (WCD, worst case distinctiveness) 作为衡量目标模糊性度量, 用来指导离线的环境重设计, 它是指智能体在目标被揭示之前可以执行的最大动作数。

目标识别设计问题可以表示为元组 $D = \langle \mathcal{P}_D, \mathcal{G}_D \rangle$, 其中 $\mathcal{P}_D$ 为问题域, $\mathcal{G}_D$ 为可能的目标集, 输出 $\mathcal{P}_D'$ 满足 $WCD(\mathcal{P}_D') \leq WCD(\mathcal{P}_D'')$ 。其中所有行动都有相应代价, 目标识别设计问题的目的是找到执行者行动的子集, 通过拦截的方式

表 2 现有目标识别设计模型属性
Table 2 Properties of existing target recognition design models

问题	文献	环境		智能体		度量		度量方法		
		部分可观	完全可观	非最优计划	部分可观	WCD	ECD	行为移除	传感器精化	行为条件
GRD	文献[64]		✓			✓		✓		
	文献[65]		✓			✓		✓		
	文献[66]		✓	✓		✓		✓		
	文献[67]	✓		✓		✓		✓	✓	
	文献[9]	✓		✓		✓		✓	✓	
	文献[68]	✓		✓		✓		✓	✓	✓
	文献[69]	✓		✓		✓		✓	✓	✓
	文献[70]		✓		✓	✓			✓	
S-GRD	文献[71]		✓			✓		✓		
	文献[72]		✓			✓	✓	✓		
	文献[73]	✓				✓		✓		

阻断执行者执行这些行动。

最初，Keren 在 2014 年提出的目标识别设计问题是在以下 3 个假设下进行的：

- ① 系统中的智能体行为将是最优的(即智能体将沿着最短的路径移动到了其目标)；
- ② 智能体是确定性的；
- ③ 环境是完全可观的(即智能体的状态和行为都是可观的)。

进一步，Keren 等于 2015 年将 GRD 问题扩展到行动部分可观测^[67]、Agent 行动非最优^[66]等情况，从而对 GRD 问题领域进行了丰富和拓展。

限制智能体可以执行的一组可用操作是改进模型的一种可行性方法。为了保持用户的舒适度，改进的模型保留所有目标的原始解决方案成本。另外，希望在最大限度地减少 WCD 的同时减少引入的更改并遵守可能指定的任何设计约束。

此外，针对现有 GRD 问题的求解过于依赖经典规划，Son 等^[65]在 2016 年提出基于答案集编程(ASP, answer set programming)描述性框架的求解方法，实验证明基于 ASP 的 GRD 问题求解不仅具有良好的扩展性，而且在效率上较传统方法能够达到 3 个数量级的升跃。

然而，许多实际应用程序必须考虑各种形式的不确定性。特别是目标识别系统通常会由于缺

少合适的传感器、传感器覆盖范围不足、传感器故障、测量结果不准确等而导致可观察性降低。在完全可观察的设置中，只有在以下情况下才会妨碍目标识别：智能体的行为可能符合多个目标，当可观察性不完整时，即使智能体的行为是特定于目标的，智能体的目标也可能无法识别。

Keren 等在 2016 年将 WCD 的概念扩展到部分可观的环境^[67]，其基础是欺骗性智能体能够控制哪些行为“发出”观测结果。在这种假设下，智能体可能保持一条模糊的路径，以获得更大的距离。同样，距离是使用最优(或“有界最优”)路径从初始状态计算的。这一策略本质上是一种蓄意欺骗，其特点是“隐私保护”。

(2) 随机性目标识别设计

在现实世界中处处存在着随机性和局限性，这导致智能体无法控制其行动的结果，同时，观察者受到传感器数量和质量的限制。

Wayllaced 等在初始的 GRD 基础上，提出了随机目标识别设计(S-GRD)^[71]，同时，提出了名为“预期情况显著性”(ECD, expected-case distinctiveness)的新指标^[72]，即根据所有目标成为真正目标的可能性来衡量可能的目标。

该框架还扩展了以下假设：智能体行为不再可观察，智能体状态只是部分可观察到的。这种

放松是以实际应用为动力的。诸如智能体导航之类的操作, 智能体操作是不可观察的, 但其状态(当前位置)可以是(至少部分)观察到的。智能体状态的部分可观测性是由于传感器分辨率低(如 GPS), 可能导致几个附近的状况彼此无法区分。最后, 将一种新类型的传感器精化看作改进某些状态可观的可能设计^[73], 同时在此基础上, 进行对传感器精化、搜索空间的剪枝和对部分可观的 S-GRD 的 ECD 度量的推广。

S-GRD 模型可以用元组 $T = \langle D, G \rangle$ 表示, 其中域 D 使用 MDP 元素, G 表示可能的目标集。计算不同的指标需要使用类似于 VI 的算法将增强的 MDP 状态和操作与可能的目标组合在一起。使用这种类型的 MDP 的原因是: 特定状态的可能目标集取决于访问该状态所到达状态的全部历史记录。随机目标识别设计使用增强 MDP 进行描述, 在元素 D 中的代价 C 只保留正代价。增强 MDP 用集合 $D = \langle \tilde{S}, \tilde{s}_0, \tilde{A}, \tilde{T}, \tilde{C} \rangle$, 每一个数组的定义如下。

状态 $\tilde{s} \in \tilde{S}$ 可以由数组 $\langle s, G' \rangle$ 表示, 其中 $s \in S$, $G' \subseteq G$ 是对于状态 s 所有可能的目标。

初始状态 $(\tilde{s}_0) = \langle s_0, G \rangle$, 行动 $\tilde{a} \in \tilde{A}$ 可以由数组 $\langle a, G' \rangle$ 表示, 其中 $a \in A$, G' 是最佳行动的所有目标的集合。

新的转换函数 $\tilde{T}: \tilde{S} \times \tilde{A} \times \tilde{S} \rightarrow [0, 1]$, 转换概率 $\tilde{T}(s, a, \tilde{s})$, 其中 $(\tilde{s}_0) = \langle s_0, G \rangle$, $(\tilde{a}_0) = \langle a_0, G' \rangle$, $\tilde{s}' = \langle s', G' \cap G \rangle$ 。

代价函数 $\tilde{C}: \tilde{S} \times \tilde{A} \times \tilde{S} \rightarrow R^+$, 代价执行行动 $\tilde{a}$ 使增加状态从 s 转换到状态 $\tilde{s}$ 。

S-GRD 算法使用增强 MDP 和 VI 类算法, 通过从增强初始状态到任何增强目标的最大期望成本来计算 WCD。

值迭代 (VI)^[74] 是找到最佳策略的基本算法之一。它使用值函数表示预期代价。初始状态 $s_0 \in S$ 的最优策略 Π^* 的预期成本是预期成本 $V(s_0)$, 所有状态 $s \in S$ 的 V 预期成本 $V(s)$ 使用贝尔曼方程计算。

$$V(s) = \min_{a \in A} \sum_{s' \in S} T(s, a, s') [C(s, a, s') + V(s')]$$

VI 的局限性在于, 即使某些状态的预期成本

已经收敛, 它也会在每次迭代中更新每个状态。拓扑值迭代 (TVI)^[75] 通过仅在一个强连接图 (SCC) 中重复更新状态直到其值收敛, 然后在另一个 SCC 中更新状态来解决此限制。由于 SCC 形成有向无环图, SCC 中的状态仅影响其之前的状态。因此, 通过以相反的拓扑排序顺序选择 SCC, 不再需要考虑状态在先前迭代中已经收敛的 SCC。

3.4.2 计划识别设计

目标识别设计是以一种易于识别智能体目标的方式设计领域的问题。而计划识别设计 (PRD, plan recognition design) 是一种可以更快地识别智能体计划的设计域问题。该问题由 Mirsky 等在 2017 正式提出^[76], 计划识别设计问题是将原 GRD 问题推广到降维区域, 使用层次计划 (GRD-PL), 该问题是利用计划库设计领域的任务, 以便于快速识别智能体的计划。GRD 可以帮助更快地解释智能体试图实现的目标, 而 PRD 则有助于更快地理解智能体如何实现其目标。在 GRD 范式的基础上, 这两个新问题 (GRD-PL 和 PRD) 都可以采用最坏情况显著性 (WCD) 进行量化。

带有计划库的目标识别设计 (GRD-PL): Ramirez 和 Geffner^[77] 提供了一种将计划库 (表示为 AND-OR 树) 编译为 STRIPS 表示的方法。从理论上讲, 可以通过解决 STRIPS 空间中的相应 GRD 问题来解决 GRD-PL 问题。但是采用该方法得到的 STRIPS 表示的复杂度可能比较高。因此, 将 GRD 问题扩展到使用计划库进行域的表示作为一种新的方法来解决这类问题。即使 WCD 计算适应于关于层次计划的推理以及如何在不限智能体实现其可能目标的情况下, 以最小化其 WCD 的方式更改给定域。在这种情况下, WCD 是使用搜索树来计算的, 该树在显示其计划之前找到智能体可以执行的最长操作序列。

计划库由数组 $\mathcal{L} = \langle \mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{R} \rangle$, 其中 $\mathcal{B}$ 是一组有限的基本动作集合, $\mathcal{C}$ 是一组有限的复杂动作集合, $\mathcal{G} \subseteq \mathcal{C}$ 是可能的目标的集合, $\mathcal{R}$ 是 $c \rightarrow \tau \mid O$ 的规则集合。其中, $c \subseteq \mathcal{C}$, τ 是 $(\mathcal{B} \cap \mathcal{C})^*$ 中的一个字符串。

$O = \{(i, j) \mid c_i \prec c_j\}$, 其中, c_i , c_j 分别代表第 i 个和第 j 个动作。

直观地讲, $\mathcal{B}$ 代表智能体可以执行的所有原子可观察动作, $\mathcal{C}$ 代表更复杂或抽象的动作, 其中 $\mathcal{G}$ 是智能体可能实现的目标, 每个 $r \in \mathcal{R}$ 表示 $\mathcal{C}$ 中的复杂动作如何分解为其他动作的序列。对于 $c_i, c_j \in \tau$, 如果存在排序约束 ($i < j \in O$), 则 $c_i \prec c_j$ 。

带有计划库的目标识别设计问题可以表示为元组 $\mathcal{D} = \langle \mathcal{P}_D, \mathcal{G}_D \rangle$, 其中 $\mathcal{P}_D$ 是由计划库表示的规划域, $\mathcal{P}_D = \langle \mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{R} \rangle$; $\mathcal{G}_D$ 是可能的目标集合; GRD-PL 问题输出的是 $\mathcal{P}'_D = \langle \mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{R}' \rangle$, 满足 $R' \subseteq R, R' \subseteq R, \mathcal{P}'_D \subseteq \mathcal{P}_D, \text{WCD}(\mathcal{P}'_D) \leq \text{WCD}(\mathcal{P}_D)$; 并且 $\mathcal{G}_D$ 中的所有目标都可以在 $\mathcal{P}_D$ 中实现。

计划识别设计: 计划识别设计是一种可以更快地识别智能体的计划的设计域问题。在 GRD 问题中, 该设计试图将所需的观察次数最小化, 直到对智能体的目标不存在任何含糊之处; 而 PRD 则试图将观察的数量降到最低, 直到对智能体的完整计划没有任何含糊之处。Mirsky 等^[78]定义了一个新的度量, 即最坏的方案区分 (WCPD)。

WCPD 的定义为: 在最坏的情况下, 能得到智能体的计划的观察数量。形式上, 对于每两个计划 p_1, p_2 , 它们的 WCPD 计算公式为

$$\text{WCPD}(p_1, p_2) = \max_{O \in (\text{OBS}(p_1) \cap \text{OBS}(p_2))} |O|$$

将定义扩展到计划库, 使计划库 $\mathcal{P}_D$ 的 WCPD 是 $\mathcal{P}_D$ 中目标生成的每一对计划中的最大 WCPD。

$$\text{WCPD}(\mathcal{P}_D) = \max_{p_1, p_2 \in \bigcup_{g \in \mathcal{G}} \text{Plans}(g)} \text{WCPD}(p_1, p_2)$$

运行示例中计划库的 WCPD 为 1, 因为在观察运行之后, 对于智能体的计划仍然存在模糊。请注意, 一个额外的观察 (如踢) 将明确地识别该智能体的计划。而计划识别设计问题就是解决计划库的 WCPD 最小化的问题。

计划识别设计可以用数组 $\mathcal{D} = \langle \mathcal{P}_D, \mathcal{L}_D \rangle$, 其中, $\mathcal{P}_D$ 是由计划库表示的规划域 $\mathcal{P}_D = \langle \mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{R} \rangle$, $\mathcal{L}_D$

是可能计划的集合 $\mathcal{L}_D = \bigcup_{g \in \mathcal{G}} \text{Plans}(g)$ 。

PRD 问题输出的是 $\mathcal{P}'_D = \langle \mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{R}' \rangle$, 满足 $R' \subseteq R, \mathcal{P}'_D \subseteq \mathcal{P}_D, \text{wcd}(\mathcal{P}'_D) \leq \text{wcd}(\mathcal{P}_D)$; 并且 $\mathcal{G}_D$ 中的所有目标都可以在 $\mathcal{P}_D$ 中实现。

GRD、PRD 与 GRD-PL 之间的比较如表 3 所示。

4 意图识别研究展望

4.1 不完美域模型

近年来, 针对不完美域模型 (imperfect domain model) 的目标识别方法^[79-80]、针对新应用场景的主动目标识别^[81]、连续空间的目标识别^[58]、非理性智能体的目标识别^[59]相继提出。除了上文中指出的问题和局限性外, 在未来研究中, 需要探索多种途径来扩展在不完美离散域模型上的目标识别工作。目前可以从以下两个设想出发进行探索: ① 使用一个传播的 RPG 来解释不完美域模型, 如文献[82]采用信息来构建规划启发式, 用于对不完美域模型进行规划; ② 文献[83]开发了一种基于规划痕迹的不完美域模型细化方法, 采用基于信息的规划痕迹和推断信息的方法获取不完美域模型。

4.2 对抗规划理论

在实际应用中, 意图识别不是最终目的。观察者的最终目的是在识别行动者真实意图的前提下, 对行动者进行协作或者阻碍。对抗条件下, 观察者可以采用更为实用的拦截措施来干扰执行者的欺骗行为, 控制目标识别过程。除了干扰对方的欺骗性规划过程, 对抗双方也能够试图控制目标识别的过程, 使规划具备可解释性 (混淆性、欺骗性、含糊性、隐私性和安全性) 以达到干扰对方的意图识别过程或者防止隐私泄露的目的。

对抗环境下的竞合关系是意图模糊与欺骗的重要因素。同时, 对手的欺骗性行为对意图识别过程的影响是未来开展意图识别的首要需求。从

表 3 GRD、PRD、GRD-PL 之间的比较
Table 3 The comparison between GRD, PRD, and GRD-PL

问题	输入	输出	度量	度量方法
GRD	STRIPS 域	带有少量行动的 STRIPS 域	WCD	明确智能体的目标所需的最大观测数
PRD	计划库	带有少量行动的计划库	WCPD	明确智能体的计划所需的最大观测数
GRD-PL	计划库	带有少量行动的计划库	WCD	明确智能体的目标所需的最大观测数

反制的角度,意图识别方法研究自然而然地引入了被识别者的欺骗性路径规划问题。目前,针对被识别者的欺骗路径规划问题,考虑其主动欺骗行为对识别效果、迟滞识别时机等方面的影响。同时,欺骗概念可以从路径规划领域延伸到一般的经典任务规划问题,如采用概率意图识别算法评估欺骗量级,并提出一个新的优化模型,在资源约束下最大化被识别者的规划欺骗性。因此,对抗环境下的意图识别与被识别者的欺骗性任务规划方法研究对无人作战、军事指挥与控制具有一定价值,也是对现有对手建模与欺骗行为建模的重要补充。

目前,一种新型的自动方法可以用来对抗 RTS 中的对手策略^[84],它结合了目标识别功能以推断对手的目标;landmark 的计算以识别可用于阻止对手达成目标的子目标;传统的自动计划以生成阻碍对手达成目标的计划。解决人工智能应用于 RTS 游戏的主要挑战之一的技术:从无到有综合计划反对对手战略。为了在实时环境中正常工作,当前的研究主要在已知域无关反规划算法的基础上进行了两项修改:① 使用代价估计而不是在目标识别中进行计划计算;② 选择任何地方的对手可以被阻挡,而不是推理最好的停止在哪里阻止它。

4.3 可解释人工智能

随着人类感知研究的新兴领域兴起,意图识别开始与另一个新兴领域——可解释性人工智能重叠^[85]。2017年,DARPA 发起“可解释人工智能”项目研究,以此探索可以使自主系统对其行为进行更好解析的技术。当前,可解释性人工智能与合法人工智能、负责任人工智能、隐私保护人工智能是可信人工智能的主要内容。针对规划决策领域可解释性研究,主要关注可解释规划、可解释性决策算法。对抗环境下,智能体的行为模型具备多模态属性,可解释行为(模型也得到了广泛关注。文献[86]对具有人类意图的智能体和人在回路中的智能体的可说明行为进行了归纳总结。考虑合作情景时,智能体的目标和规划必须同时具备明确性、透明性、解释性和预测性^[87]。考虑对抗情景时,智能体的目标和规划必须具备混淆性、含糊性、隐私性和安全性。近年来,为

实现多模态的规划决策方法,控制信息披露和隐藏的规划方法、控制目标混淆或明确的规划方法和控制意图分享与隐藏的决策方法相继被提出。一些新的研究包括将意图识别应用于多智能体行为识别中^[6]、决策过程中控制意图的分享与隐藏^[7]。这些与可解释性规划^[8]、隐私保护规划密切相关^[7,9],其中任务执行者可根据相关情景主动披露或隐秘混淆其真实目标。

4.4 多智能体计划识别

在复杂的战场环境中,智能体之间通常采用协同的方式实现共同的目标,即联合意图。考虑到一个智能体的行为对与其协同作战的智能体产生的影响,针对多智能体的联合意图识别比单个智能体的意图复杂和困难更加符合实际的作战需求。多智能体意图识别的重点和难点是基于统一的建模框架来描述智能体的协同行为。如果识别者具备领域背景和识别对象行为习惯的相关知识,那么通常可使用显示建模的方式,即采用一个复杂的有向图模型对各要素间的依赖关系进行详细建模。

多智能体计划识别(MAPR)^[83,88-89]是计划、活动和意图识别(PAIR)研究主题的一个子集^[1],MAPR 研究的重点是观察单个智能体的行为,并从这些行为中推断哪些团队中的智能体进行了工作,以及团队的目的^[90]。目前大多数 MAPR 解决方案是识别特定领域的活动,依赖于将观测结果与人类生成的库相匹配,并且预先分析完整结构的同步痕迹。

5 实际应用

5.1 军事辅助决策

CADET 系统是一个集中在旅级地面行动的战斗规划系统,其中重要一环就是对抗推理和对抗规划。DARPA 从 2004 年开始筹建实时对抗情报和决策(RAID, real-time adversarial intelligence and decision-making)计划^[91],关注如何将敌人在军事行动中的对抗行为纳入规划考虑范畴。

RAID 系统面临许多问题,红蓝双方的行动有很强的相互依赖关系,一方对另一方的资产和行动知识不可避免地受到限制。观察和观察的解释都有一定限度的误差和错误。除了局部的、延迟

的甚至是错误的观察外,对于战场的知识也会受到有目的性的、持续的、攻击性的、智能的隐藏和欺骗。由于文化上、条令条例上和心理上的影响,作战规划时往往仅仅考虑最危险(通常是敌方理论上最优的)的作战行动序列。然而实际最可能发生的作战行动序列很大程度上会受到人为因素的影响,甚至可能与理论上的最危险策略不同。以城市战斗为例,复杂的城市地理环境使战斗呈现高密度同时有碎片化的威胁和机遇^[9];地形由于人的参与也是动态变化的;同时战场上的非战斗人员必须明确纳入考虑范围,并最小化其附加损害。除了火力和机动需要考虑之外,其他的如情报收集、后勤等都与火力机动紧密耦合,这些都充满着对抗性和不确定性。

5.2 即时战略游戏

即时战略(RTS, real-time strategy)游戏是一种对复杂战场环境模拟的仿真手段,在仿真过程中,己方人员指示己方实体通过获取资源、构造结构,来摧毁对手的建筑从而赢得比赛。RTS游戏是对作战仿真的简化,因此,可以作为即时对抗规划、不确定条件下的决策支持等方法的试验床。由于庞大的状态空间、动态的对抗环境,以及对决策时效性的高度要求,RTS游戏的研究使AI研究人员面临极大的挑战。尤其是在对指挥员智能体建模时,需要考虑到全局信息。在广泛的研究范围中,作为实时对抗性规划和不确定性决策的基本问题,人工智能规划已经成为一个至关重要的研究领域。

RTS游戏由于存在许多具有挑战性的子问题,如目标推理、策略构建、学习等,已被广泛应用于AI研究。应用于RTS游戏的AI的主要挑战之一是自主开发、综合考虑对手观察结果的计划。该任务可以看作对策问题。

5.3 对抗规划

对抗规划划分为对抗推理和对抗计划两个相互影响的部分,来表示在一个对抗环境下,一方通过计算求解对手的决定性状态、意图和行为并尽全力反制敌方的行动和计划的动态规划过程。这个领域的子问题包括信念和意图识别、对手策略预测、规划识别、欺骗计划发现、欺骗计划和计划生成等。从工程角度来讲,对抗计划的研究

覆盖了广泛的实际问题领域,除了即时战略游戏这类娱乐性策略游戏之外,还可以应用在军事计划和指挥、军事和外交情报获取、反恐和国土安全、信息安全、仿真和训练系统、实用机器人等领域。

目前,规划算法应用于机器人、航空航天和工业工程等领域,其目的是找到困难问题的优化解决方案。这些问题可能涉及数千个参数和相互作用,不可能通过手工获得一个解决方案。例如,考虑一个包含数千台具有大量数据的服务器的分布式计算基础设施,规划器将被用来安排数据迁移、启动新的实例、处理数据、以最有效的方式分配资源。通常,目标是在所有服务器上有一个相当分布的工作负载。显然,找到最优的计划(即成本最低的计划)并不是一件容易的事情。

规划器是确定性的,意味着可以产生相同的输入(目标、初始状态等)。对手可以从这些输入—输出对进行学习,进而得到行为模型。这样对手可以在给定输入时,以高概率预测出输出计划中的关键动作。当计划人员针对分布式计算调度问题执行任务时,对手可以推断出计划的哪一步至关重要。即如果避免了那一步,资源或计算位置将产生最大的负面影响。这意味着对手可以识别出最感兴趣的服务器或攻击的链接,并且采取了相应的措施。

如今,通常智能体在某处想要使用计划系统找到自己的位置和自己的路线是一件很容易的事(应用GPS)。路径规划器用于找到从两个(或多个)点导航的最佳方法,其根据每个路径片段的代价(即速度限制、路程等)输出一组方向。考虑在大规模使用计划系统而没有经过人工验证的情况下应用的所有问题(机场、数据中心、生产线等)。确定计划算法可能使用的关键目标,可以提供发动攻击的手段,从而增加所需的资源(时间、计算能力、能源等)或阻止规划器找到任何合理的计划。

5.4 人机协同

当前,许多智能系统使用固定通信输入或预设响应中至少一个与其他系统进行交互,从而导致僵化的交互体验以及为系统开发各种场景的大量努力。固定输入限制了用户的自然行为,从而

限制了沟通,并且预设响应会阻止系统适应当前情况,除非专门实施。相反,闭环交互专注于动态响应,这些动态响应基于对用户感知活动的解释来说明用户当前正在执行的操作。采用闭环交互的智能体还可以监视其交互,以确保用户按预期进行响应。通过演示实现了一个辅助交互式智能体,该智能体集成了计划、计划识别和意图识别,以预测用户要完成的任务并自主决定响应这些预测要采取的措施。

当交互式体验开始时,辅助智能体将没有用户模型。这也意味着智能体不知道用户想要做什么,必须通过观察才能做出明智的决定。用户将在模拟环境中玩回合制游戏,可以自由地从一组完成标准中进行选择,成功满足其中任何一个条件都将赢得比赛。用户将在第一个回合中采取行动,这为智能体提供一些有关它们打算完成哪些事件的信息。当智能体收到足够的信息来决定如何做出响应,则智能体将在第二轮行动。由于演示的模拟游戏设置,用户输入则为简单的离散按钮按下和鼠标点击,不需要任何原始传感器数据。因此,这种实现只是将活动识别作为从输入到游戏相应动作的映射来执行。计划和意图识别组件收到这些操作后,通过观察概率识别计划(R&G方法)^[37]进行识别,该算法运行生成规划器来模拟用户解决各种问题。

6 结束语

本文对手建模进行了简要的介绍,并针对行为建模中的目标识别、计划识别、目标识别设计、计划识别设计共4方面现有的研究方法进行了详细的总结与分析;同时,给出在不完美域模型、对抗规划、意图识别控制与可解释人工智能、多智能体意图识别等方面在意图识别领域的研究展望;针对军事辅助决策、即时战略游戏、对抗规划、人机协同的典型应用分析,指出了这一研究工作的实际意义。

尽管Keren等在最新的综述中总结了当前目标识别设计方面的主要工作^[92],现有的研究工作极大地提升了目标识别的速度与正确率,但这些工作仍然是不完善的,仍然有新的应用领域以及改进方向为研究工作提出挑战。这些挑战多数来

自现有的有限计算能力、人类行为的不确定性以及现实场景中的特殊性质。

参考文献:

- [1] SUKTHANKAR G. Plan, activity, and intent recognition: theory and practice[R]. 2014.
- [2] CHAKRABORTI T, KAMBHAMPATI S, SCHEUTZ M, et al. AI challenges in human-robot cognitive teaming[J]. arXiv preprint arXiv:1707.04775, 2017.
- [3] ALBRECHT S V, STONE P. Autonomous agents modelling other agents: a comprehensive survey and open problems[J]. Artificial Intelligence, 2018, 258: 66-95.
- [4] HEINZE C. Modelling intention recognition for intelligent agent systems[R]. 2004.
- [5] LE GUILLARME N. A game-theoretic planning framework for intentional threat assessment[D]. Thèse de doctorat: Université de Caen, 2016.
- [6] BIGELOW D. Intent recognition in multi-agent domains[M]. University of Nevada, Reno, 2013.
- [7] STROUSE D J, KLEIMAN-WEINER M, TENENBAUM J, et al. Learning to share and hide intentions using information regularization[C]//Advances in Neural Information Processing Systems. 2018: 10249-10259.
- [8] CHAKRABORTI T, KULKARNI A, SREEDHARAN S, et al. Explicability legibility predictability transparency privacy security the emerging landscape of interpretable agent behavior[C]//Proceedings of the International Conference on Automated Planning and Scheduling. 2019: 86-96.
- [9] KEREN S, GAL A, KARPAS E. Privacy preserving plans in partially observable environments[C]//IJCAI. 2016: 3170-3176.
- [10] WRIGHT J R. Modeling human behavior in strategic settings[D]. Columbia: University of British Columbia, 2016.
- [11] PLONSKY O, APEL R, ERT E, et al. Predicting human decisions with behavioral theories and machine learning[J]. arXiv preprint arXiv:1904.06866, 2019.
- [12] BORGHETTI B J. Opponent modeling in interesting adversarial environments[M]. Minnesota: University of Minnesota, 2008.
- [13] BROWNE C B, POWLEY E, WHITEHOUSE D, et al. A survey of monte carlo tree search methods[J]. IEEE Transactions on Computational Intelligence and AI in Games, 2012, 4(1): 1-43.
- [14] CHAKRABORTY D, STONE P. Multiagent learning in the presence of memory-bounded agents[J]. Autonomous Agents and Multi-Agent Systems, 2014, 28(2): 182-213.
- [15] KOLODNER J. Case-based reasoning[M]. Morgan Kaufmann, 2014.
- [16] CARMEL D, MARKOVITCH S. Learning models of intelligent agents[C]//AAAI/IAAI. 1996: 62-67.
- [17] BAARSLAG T, HENDRIKX M J C, HINDRIKS K V, et al. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques[J]. Autonomous Agents and Multi-Agent Systems, 2016, 30(5): 849-898.

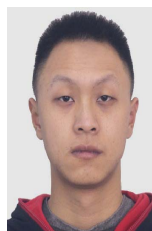
- [18] BARRETT S, STONE P, KRAUS S, ET al. Teamwork with limited knowledge of teammates[C]//Twenty-Seventh AAAI Conference on Artificial Intelligence. 2013.
- [19] ALBRECHT S V, CRANDALL J W, RAMAMOORTHY S. An empirical study on the practical impact of prior beliefs over policy types[C]//Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.
- [20] ALBRECHT S V, RAMAMOORTHY S. On convergence and optimality of best-response learning with policy types in multiagent systems[J]. arXiv preprint arXiv:1907.06995, 2019.
- [21] SCHADD F, BAKKES S, SPRONCK P. Opponent modeling in real-time strategy games[C]//GAMEON. 2007: 61-70.
- [22] WEN Y, YANG Y, LU R, et al. Multi-agent generalized recursive reasoning[J]. arXiv preprint arXiv:1901.09216, 2019.
- [23] WEN Y, YANG Y, LUO R, et al. Probabilistic recursive reasoning for multi-agent reinforcement learning[J]. arXiv preprint arXiv:1901.09207, 2019.
- [24] DOSHI P, ZENG Y, CHEN Q. Graphical models for interactive POMDPs: representations and solutions[J]. Autonomous Agents and Multi-Agent Systems, 2009, 18(3): 376.
- [25] TORKAMAN A, SAFABAKHSH R. Robust opponent modeling in real-time strategy games using bayesian networks[J]. Journal of AI and Data Mining, 2019, 7(1): 149-159.
- [26] MAO W, GRATCH J, LI X. Probabilistic plan inference for group behavior prediction[J]. IEEE Intelligent Systems, 2012, 27(4): 27-36.
- [27] HAUSKNECHT M, MUPPARAJU P, SUBRAMANIAN S, et al. Half field offense: an environment for multiagent learning and ad hoc teamwork[C]//AAMAS Adaptive Learning Agents (ALA) Workshop. 2016.
- [28] ŠOŠIĆ A. Learning models of behavior from demonstration and through interaction[D]. Technische Universität, 2018.
- [29] HERNANDEZ-LEAL P, ZHAN Y, TAYLOR M E, et al. Efficiently detecting switches against non-stationary opponents[J]. Autonomous Agents and Multi-Agent Systems, 2017, 31(4): 767-789.
- [30] ALBRECHT S V, RAMAMOORTHY S. Are you doing what i think you are doing? criticising uncertain agent models[J]. arXiv preprint arXiv:1907.01912, 2019.
- [31] WANG Z, BOULARIAS A, MÜLLING K, et al. Balancing safety and exploitability in opponent modeling[C]//Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011.
- [32] STANESCU A M. Outcome prediction and hierarchical models in real-time strategy games[R]. 2019.
- [33] MOURAD M, AREF M, ABD-ELAZIZ M. Opponent models pre-processing in real-time strategy games[J]. International Journal of Intelligent Computing and Information Sciences, 2016, 16(3): 37-45.
- [34] SUKTHANKAR G R. Activity recognition for agent teams[R]. 2007.
- [35] FREEDMAN R G, ZILBERSTEIN S. A unifying perspective of plan, activity, and intent recognition[C]//Proceedings of the AAAI Workshops: Plan, Activity, Internet Recognition. 2019: 1-8.
- [36] RAMÍREZ M, GEFFNER H. Plan recognition as planning[C]//Twenty-First International Joint Conference on Artificial Intelligence. 2009.
- [37] SOHRABI S, RIABOV A V, UDREA O. Plan recognition as planning revisited[C]//IJCAI. 2016: 3258-3264.
- [38] PEREIRA R F, OREN N, MENEGUZZI F. Landmark-based approaches for goal recognition as planning[J]. arXiv preprint arXiv:1904.11739, 2019.
- [39] AINETO D, JIMÉNEZ S, ONAINDIA E, et al. Model recognition as planning[C]//Proceedings of the International Conference on Automated Planning and Scheduling. 2019: 13-21.
- [40] ANG S, CHAN H, JIANG A X, et al. Game-theoretic goal recognition models with applications to security domains[C]//International Conference on Decision and Game Theory for Security. 2017: 256-272.
- [41] LE GUILLARME N, MOUADDIB A I, LEROUVREUR X, et al. A generative game-theoretic framework for adversarial plan recognition[C]//JFPDA 2015. 2015.
- [42] LI J, REN T, SU H, et al. Learn a robust policy in adversarial games via playing with an expert opponent[C]//Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems. 2019: 2096-2098.
- [43] ŠOŠIĆ A. Learning models of behavior from demonstration and through interaction[D]. Technische Universität, 2018.
- [44] ZIEBART B D. Modeling purposeful adaptive behavior with the principle of maximum causal entropy[D]. Figshare, 2010.
- [45] TASTAN B. Learning human motion models[C]//Eighth Artificial Intelligence and Interactive Digital Entertainment Conference. 2012.
- [46] GAURAV S, ZIEBART B. Discriminatively learning inverse optimal control models for predicting human intentions[C]//Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. 2019: 1368-1376.
- [47] LI X, YANG W, ZHANG Z. A unified framework for regularized reinforcement learning[J]. arXiv preprint arXiv:1903.00725, 2019.
- [48] TIAN Z, WEN Y, GONG Z, et al. A regularized opponent model with maximum entropy objective[J]. arXiv preprint arXiv:1905.08087, 2019.
- [49] MAYNARD M, DUHAMEL T, KABANZA F. Cost-based goal recognition meets deep learning[J]. arXiv preprint arXiv:1911.10074, 2019.
- [50] WOOKHEE M, YOUNG E H, ROWE J. Deep learning-based goal recognition in open-ended digital games[C]//Tenth AAAI Conference on Artificial Intelligence & Interactive Digital Entertainment. 2014.
- [51] DUHAMEL T, MAYNARD M, KABANZA F. A transfer learning method for goal recognition exploiting cross-domain spatial features[J]. arXiv preprint arXiv:1911.10134, 2019.
- [52] THIBAUT D, MAYNARD M, KABANZA F. Imagination-augmented deep learning for goal recognition[J]. arXiv preprint arXiv:2003.09529v1, 2020.
- [53] BLAYLOCK N, ALLEN J. Fast hierarchical goal schema recognition[C]//Proceedings of the National Conference on Artificial Intelligence. 2006: 796.

- [54] VERED M, KAMINKA G A. Heuristic online goal recognition in continuous domains[C]//International Joint Conference on Artificial Intelligence. 2017: 4447-4454.
- [55] VERED M, KAMINKA G A, BIHAM S. Online goal recognition through mirroring: Humans and agents[C]//The Fourth Annual Conference on Advances in Cognitive Systems. 2016.
- [56] VERED M, KAMINKA G A. Online recognition of navigation goals through goal mirroring[C]//Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems. International Foundation for Autonomous Agents and Multiagent Systems. 2017: 1748-1750.
- [57] VERED M, PEREIRA R F, MAGNAGUAGNO M C, et al. Towards online goal recognition combining goal mirroring and landmarks[C]//AAMAS. 2018: 2112-2114.
- [58] MASTERS P, SARDINA S. Cost-based goal recognition for the path-planning domain[C]//IJCAI. 2018: 5329-5333.
- [59] MASTERS P, SARDINA S. Goal recognition for rational and irrational agents[C]//Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems. 2019: 440-448.
- [60] HOFFMANN J, PORTEOUS J, SEBASTIA L. Ordered landmarks in planning[J]. Journal of Artificial Intelligence Research, 2004, 22: 215-278.
- [61] SCHMIDT C, SRIDHARAN N, GOODSON J. The plan recognition problem: an intersection of psychology and artificial intelligence[J]. Artificial Intelligence, 1978, 11: 45-83.
- [62] COHEN P R, PERRAULT C R, ALLEN J F. Beyond question answering[M]//Strategies for Natural Language Processing. Lawrence Erlbaum Associates, 1981.
- [63] PENTNEY W, POPESCU A, WANG S, KAUTZ H, et al. Sensor-based understanding of daily life via large-scale use of common sense[C]//Proceedings of AAAI. 2006.
- [64] KEREN S, GAL A, KARPAS E. Goal recognition design[C]//Twenty-Fourth International Conference on Automated Planning and Scheduling. 2014.
- [65] SON T C, SABUNCU O, Schulz-Hanke C, et al. Solving goal recognition design using ASP[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [66] KEREN S, GAL A, KARPAS E, et al. Goal recognition design for non-optimal agents[C]//National Conference on Artificial Intelligence. 2015: 3298-3304.
- [67] KEREN S, GAL A, KARPAS E. Goal recognition design with non-observable actions[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [68] KEREN S, GAL A, KARPAS E. Strong stubborn sets for efficient goal recognition design[C]//Twenty-Eighth International Conference on Automated Planning and Scheduling. 2018.
- [69] SARAH K, AVIGDOR G, EREZ K. Goal recognition design in deterministic environments[J]. Journal of Artificial Intelligence Research, 2019, 65: 209-269.
- [70] KEREN S, PINEDAL, GAL A, et al. Equi-reward utility maximizing design in stochastic environments[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017: 4353-4360.
- [71] WAYLLACE C, HOU P, YEOH W, et al. Goal recognition design with stochastic agent action outcomes[C]//IJCAI. 2016.
- [72] WAYLLACE C, HOU P, YEOH W. New Metrics and Algorithms for Stochastic Goal Recognition Design Problems[C]//IJCAI. 2017: 4455-4462.
- [73] WAYLLACE C, KEREN S, YEOH W, et al. Accounting for partial observability in stochastic goal recognition design: messing with the marauder's map[C]//Proceedings of the 10th Workshop on Heuristics and Search for Domain-Independent Planning (HSDIP), Delft, The Netherlands. 2018: 33-41.
- [74] RICHARD B. Dynamic Programming[M]. Princeton University Press, 1957.
- [75] ROBERT T. Depth-first search and linear graph algorithms[J]. SIAM Journal on Computing, 1972, 1(2): 146-160.
- [76] MIRSKY R, STERN R, GAL Y, et al. Plan recognition design[C]//Workshops at the Thirty First AAAI Conference on Artificial Intelligence. 2017.
- [77] RAMIREZ M, GEFFNER H. Heuristics for planning, plan recognition and parsing[J]. arXiv preprint arXiv:1605.05807, 2016.
- [78] MIRSKY R. Goal and plan recognition design for plan libraries[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2): 14.
- [79] PEREIRA R F, PEREIRA A G, MENEGUZZI F. Landmark-enhanced heuristics for goal recognition in incomplete domain models[C]//Proceedings of the International Conference on Automated Planning and Scheduling. 2019: 329-337.
- [80] PEREIRA R F. Goal recognition over imperfect domain models[J]. arXiv preprint arXiv:2005.05712, 2020.
- [81] AMATO C, BAISERO A. Active goal recognition[J]. arXiv preprint arXiv:1909.11173, 2019.
- [82] ZHANG T. Solving large scale linear prediction problems using stochastic gradient descent algorithms[C]//Proceedings of the International Conference on Machine Learning (ICML). 2004. 919-926.
- [83] ZHUO H H. Recognizing multi-agent plans when action models and team plans are both incomplete[J]//ACM Transactions on Intelligent Systems and Technology, 2019, 10(3): 1-24.
- [84] POZANCO A, MARTIN Y E, FERNANDEZ S, et al. Counterplanning using Goal Recognition and Landmarks[C]//International Joint Conference on Artificial Intelligence. 2018: 4808-4814.
- [85] GADEPALLY V, GOODWIN J, KEPNER J, et al. AI enabling technologies: a survey[J]. arXiv preprint arXiv:1905.03592, 2019.
- [86] CHAKRABORTI T, KULKARNI A, SREEDHARAN S, et al. Explicability legibility predictability transparency privacy security the emerging landscape of interpretable agent behavior[C]//Proceedings of the International Conference on Automated Planning and Scheduling. 2019: 86-96.
- [87] SREEDHARAN S, KAMBHAMPATI S. Balancing explicability and explanation in human-aware planning[C]//2017 AAAI Fall Symposium Series. 2017.

- [88] BANERJEE B, KRAEMER L, LYLE J, et al. Multi-agent plan recognition: formalization and algorithms[C]//National Conference on Artificial Intelligence, 2010: 1059-1064.
- [89] ZHUO H H. Multiagent plan recognition from partially observed team traces[J]. Plan, Activity, and Intent Recognition, 2014: 227-249.
- [90] ARGENTA C, DOYLE J. Multi-agent plan recognition as planning (MAPRAP)[C]//International Conference on Agents and Artificial Intelligence. 2016: 141-14
- [91] ALEXANDER K, WILLIAM M. McEneaney. Adversarial reasoning: computational approaches to reading the opponent's mind[R]. 2006.
- [92] KEREN S, Gal A, KARPAS E. Goal recognition design-survey[C]//Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI-20). 2020.



罗俊仁（1989- ），男，湖北大冶人，国防科技大学博士生，主要研究方向为智能体建模、对抗团队博弈、多智能体强化学习。



袁唯淋（1994- ），男，云南曲靖人，国防科技大学博士生，主要研究方向为安全博弈、对手建模、强化学习、多智能体系统。

[作者简介]



高巍（1996- ），女，辽宁开原人，国防科技大学硕士生，主要研究方向为对手建模、任务规划、意图识别、弹道规划。



张万鹏（1981- ），男，四川邛崃人，国防科技大学副研究员，主要研究方向为智能决策、任务规划、自动化和控制、人机协同。