北京大学

博士研究生学位论文

题目：多智能体环境中面向决策的对手建模算法研究

姓　　名：于晓彭

学　　号：2201111706

院　　系：计算机学院

专　　业：计算机应用技术

研究方向：多媒体智能计算与机器学习

导　　师：卢宗青 长聘副教授

■ 学术学位　　□ 专业学位

二〇二六年五月

摘要

多智能体强化学习是人工智能领域的重要研究方向，近年来在博弈对抗、协同控制、资源调度、自动驾驶编队及复杂社会模拟等任务中展现出广阔应用前景。与单智能体强化学习不同，多智能体系统中多个决策实体在共享环境中持续交互，任一智能体的行为不仅影响环境状态，还会反馈到其他参与者，从而使环境对各个智能体呈现显著的非平稳性。这种非平稳性破坏了传统强化学习关于环境转移概率固定与策略稳定的基本假设，使标准学习算法在多智能体场景中面临收敛困难、训练不稳定以及泛化能力不足等问题。从博弈论视角看，多智能体交互本质上是动态博弈过程，智能体不仅需要适应环境变化，还必须应对其他参与者不断演化的策略行为，因此，对手建模成为缓解环境非平稳性、提升决策性能的重要途径。通过建模其他参与者的策略模式、高层意图乃至学习动态，智能体能够形成更强的社会智能，并据此优化交互策略，这对非合作博弈中的竞争决策以及协作任务中的效率提升都具有重要意义。

随着应用场景逐渐向开放性与异质性发展，对手策略往往事先未知且持续演化，这对对手模型的表达能力、适应能力以及在线更新能力提出了更高要求。从方法论角度来看，对手建模面临三个相互关联的核心挑战：表示层，如何选择合适的抽象层级刻画对手行为；推理层，如何在部分可观测环境中对对手隐藏信息进行有效推断；决策集成层，如何将对手模型高效融入规划与决策过程。围绕如何在多智能体环境中构建高效、结构化且可适应的对手建模方法这一核心科学问题，本文从表示建模、推理机制与决策集成三个互补视角展开研究，并提出了三种具有代表性的对手建模方法。

首先，在表示层，针对对手长期行为结构难以刻画的问题，提出一种基于子目标表示的对手建模方法（OMG）。该方法将对手建模从传统的低层动作拟合提升至高层意图建模层面，通过引入显式子目标作为中间变量刻画对手行为的内在结构。具体而言，利用变分推断模型提取对手的子目标分布，并设计子目标推断器与选择器分离训练机制，使高层意图与低层决策形成紧密耦合。实验结果表明，在稀疏奖励与目标导向博弈任务中，该方法相比传统基于动作的建模方式表现出更快的学习速度与更强的泛化能力。

其次，在推理层，针对信息不完全与多对手交互场景中语义推理复杂的问题，提出一种基于显式反馈优化的多对手推理建模方法（EMO）。该方法为每个对手构建独立的原子对手模型，并引入中央验证器构成双层优化结构，实现多对手推理结果的一致性校验与协同修正。通过显式建模对手的私有状态与推理过程，该方法将对手建模从隐式潜变量推断提升为可解释的显式推理结构。实验结果显示，该方法在典型不完

全信息博弈任务中显著提升了隐藏信息推断准确率与多轮推理一致性。

再次，在决策集成层，针对对手模型难以高效嵌入规划过程的问题，提出一种结合策略类型建模与信念感知规划的对手建模方法（QOM）。该方法通过量化自编码器将复杂连续的对手策略空间压缩为有限离散类型集合，并基于贝叶斯滤波机制进行在线信念跟踪。在此基础上，将信念加权元策略与蒙特卡洛树搜索相结合，实现对动态对手行为的自适应规划。理论分析表明，在真实对手属于固定策略类型的条件下，该方法能够获得次线性遗憾上界。实验结果显示，在合作—对抗混合博弈任务中，该方法能够有效识别对手策略类型并做出鲁棒响应。

综上，围绕表示建模、显式推理与决策集成三个互补视角，系统研究了多智能体环境中的对手建模问题。OMG 侧重于对手行为结构的表示学习，EMO 侧重于不完全信息场景下的显式推理建模，QOM 则侧重于对手模型与规划决策过程的有效结合。三项工作分别从表示、推理与决策集成三个层面拓展了对手建模方法，为复杂多智能体环境中的智能决策提供了新的技术思路。理论分析与实验评估表明，所提出的方法在多个代表性多智能体测试环境中取得了优于现有对手建模方法的表现，并在样本效率、策略适应性、泛化能力以及对未知行为模式的响应能力等方面展现出综合优势。

关键词：对手建模，策略学习，强化学习，多智能体强化学习，决策规划

Decision-Oriented Opponent Modeling in Multi-Agent Environments

Xiaopeng Yu (Computer Application Technology)
Advised by Zongqing Lu Tenured Associate Professor

ABSTRACT

Multi-agent reinforcement learning is an important research direction in artificial intelligence and has shown broad application prospects in tasks such as game playing, cooperative control, resource scheduling, autonomous driving formation, and complex social simulation. Different from single-agent reinforcement learning, multi-agent systems involve multiple decision-making entities that interact continuously in a shared environment. The behavior of each agent not only affects the environment state but also influences other participants, making the environment highly non-stationary from the perspective of each agent. This non-stationarity breaks the basic assumptions of traditional reinforcement learning regarding fixed transition dynamics and stable policies, causing standard learning algorithms to suffer from convergence difficulties, training instability, and insufficient generalization in multi-agent settings. From the perspective of game theory, multi-agent interaction is essentially a dynamic game process, in which agents must not only adapt to environmental changes but also respond to the evolving strategies of other participants. Therefore, opponent modeling has become an important approach for mitigating environmental non-stationarity and improving decision-making performance. By modeling the strategy patterns, high-level intentions, and even learning dynamics of other participants, an agent can acquire stronger social intelligence and optimize its interactive policy accordingly, which is important for both competitive decision-making in non-cooperative games and efficiency improvement in cooperative tasks.

As application scenarios become increasingly open and heterogeneous, opponents' strategies are often unknown in advance and may evolve continuously, which places higher demands on the expressive power, adaptability, and online updating capability of opponent models. From a methodological perspective, opponent modeling faces three closely related core challenges: the representation level, namely how to choose an appropriate abstraction level to characterize opponent behaviors; the reasoning level, namely how to effectively infer hidden

information of opponents under partially observable environments; and the decision integration level, namely how to efficiently incorporate opponent models into planning and decision-making processes. To address the core scientific problem of how to construct efficient, structured, and adaptive opponent modeling methods in multi-agent environments, this dissertation investigates the problem from three complementary perspectives: representation modeling, reasoning mechanisms, and decision integration, and proposes three representative opponent modeling methods.

First, at the representation level, an opponent modeling framework based on subgoal representation, termed OMG, is proposed to address the difficulty of characterizing the long-term behavioral structure of opponents. This method elevates opponent modeling from traditional low-level action fitting to high-level intention modeling by introducing explicit subgoals as intermediate variables to capture the intrinsic structure of opponent behaviors. Specifically, a variational inference model is used to extract the subgoal distribution of opponents, and a decoupled training mechanism for the subgoal inference module and the subgoal selector is designed to tightly couple high-level intention inference with low-level decision making. Experimental results show that, in sparse-reward and goal-oriented game tasks, this method achieves faster learning and stronger generalization than conventional action-based opponent modeling methods.

Second, at the reasoning level, an explicit feedback optimization based multi-opponent reasoning method, termed EMO, is proposed to address the complexity of semantic reasoning in scenarios with incomplete information and multiple interacting opponents. This method constructs an independent atomic opponent model for each opponent and introduces a central verifier to form a two-level optimization structure, enabling consistency checking and collaborative correction of multi-opponent reasoning results. By explicitly modeling the private states and reasoning processes of opponents, this method upgrades opponent modeling from implicit latent-variable inference to an interpretable explicit reasoning structure. Experimental results demonstrate that, in typical imperfect-information game tasks, this method significantly improves the accuracy of hidden information inference and the consistency of multi-round reasoning.

Third, at the decision integration level, a method termed QOM is proposed to address the difficulty of efficiently incorporating opponent models into planning processes. QOM combines strategy-type modeling with belief-aware planning. It compresses the complex continuous strategy space of opponents into a finite set of discrete strategy types through a quantized

autoencoder, and performs online belief tracking based on a Bayesian filtering mechanism. On this basis, a belief-weighted meta-policy is integrated with Monte Carlo Tree Search to achieve adaptive planning against dynamic opponents. Theoretical analysis shows that, when the true opponent belongs to a fixed strategy type, the method achieves a sublinear regret bound. Experimental results further show that, in mixed cooperative-competitive game tasks, QOM can effectively identify opponent strategy types and produce robust responses.

In summary, this dissertation systematically studies opponent modeling in multi-agent environments from three complementary perspectives: behavioral structure modeling, natural language reasoning, and strategy-type abstraction. OMG focuses on representation learning of opponent behavioral structure, EMO emphasizes explicit reasoning under imperfect-information settings, and QOM centers on the effective integration of opponent models with planning and decision making. These three methods extend opponent modeling from the levels of representation, reasoning, and decision integration, respectively, and provide new technical insights for intelligent decision making in complex multi-agent environments. Theoretical analysis and experimental evaluation show that the proposed methods achieve better performance than existing opponent modeling approaches in multiple representative multi-agent benchmark environments, and demonstrate comprehensive advantages in sample efficiency, strategy adaptability, generalization ability, and responsiveness to unknown behavior patterns.

KEYWORDS: Opponent Modeling, Strategy Learning, Reinforcement Learning, Multi-Agent Reinforcement Learning, Decision Planning