Review of Multi-Agent

智能体BDI

Belief: 对环境和自身的评价

Desire: 实体希望达到的事件状态。目标，不与动作直接相关，只指明方向

Intention: 实体决定实施动作的事件状态。与动作直接相关，决定实施动作

动态转化过程：信念修正、愿望产生、意图生成、动作选择

多智能体系统

分而治之、相互协作

分类：集中式、分布式、混合式

通信方式：直接发送、广播、黑板、公告板

冲突形式：任务冲突、空间冲突、信息冲突

合同网协议：委托和承揽关系

结构：本地数据库、通信处理器-网络、合同处理器、任务处理器

信息融合

检测、相关、组合、估计；位置估计、身份估计精度；环境态势评估和威胁估计

要点

硬件基础：多传感器系统

加工对象：多源信息

核心：协调优化和综合处理

结构

检测级：并行、串行、树状

位置级：集中、分布、混合

属性~：数据层、特征层、决策层

OODA

Observe Orient Decide Action

Agent体系结构

慎思型

把Agent看作知识系统，符号AI的方法，BDI的形式化描述，逻辑与推理。智能高但执行慢

反应型

感知和行动，“感知-动作”模型。反应快，快速适应环境变化，智能低

混合型

高层：认知层，符号AI，规划和决策

低层：反应层，快速响应突发事件，优先级更高

Markov博弈

单智能体Markov决策，多智能体Markov博弈（随机博弈）

以纳什均衡作为协作的目标

极大极小Q

双人（i, -i）零和博弈

Nash Q

多人（1, 2, 3, ..., n）一般和随机博弈

CTDE 集中训练分散执行

中央控制器辅助训练，执行时独立根据本地观测做决策。实时决策。

优点：有利于学到更好的策略。而去中心化决策无需通信，可以做到实时决策。

基于值分解的多智能体强化学习方法

用于团队合作。将团队Q分解到个体。算法：VDN, QMIX, QTRAN

需要解决的问题：

1. 信用分配。评价智能体策略对团队的贡献
2. 虚假奖励。高的团队奖励可能是队友导致的。
3. 懒惰智能体。部分智能体学到了比较好的策略且能完成任务。其他智能体不需要再做什么。

混合增强智能

包括人在回路和认知计算

1. 人在回路。需要与人进行交互的智能模型系统。人始终是系统的一部分。
2. 认知计算。模仿人脑功能并提高计算机推理、决策、感知能力的软硬件。

建立关于大脑感知、推理、响应刺激的模型。因果模型、直觉推理模型、联想记忆模型。

人在回路

定义

人对模糊问题、不确定问题的高级认知机制与机器智能系统紧密耦合。人的感知、认知能力和计算机强大的存储、运算能力相结合，1+1>2

框架

机器学习模型做判断，置信度低时人类介入。人类数据作为训练数据继续训练模型，新的知识记录到数据库。

研究内容

1. 如何用自然的方式训练机器，突破人机交互屏障
2. 如何将人类与机器的优势相结合，实现高效的人机协同构建
3. 如何构建跨任务、跨领域的上下文关系
4. 如何建立任务或概念驱动的机器学习方法，使机器能够从人类知识中学习。

认知计算

框架

感知、注意、理解、证实、规划、评测

自上而下的选择性注意：基于规划

自下而上的选择性注意：基于感知

基于理解或规划的评测：先验概率（预测）

基于感知的评测：后验概率（实测）

认知计算的过程：根据满足目标所需要的信息与外界不断交互，逐渐将事物展开的思维活动，而非简单的基于知识的处理。强调不断交互和展开。

证实：下一步该做什么？是否达到预期？继续还是尝试其他方法？

直觉

人脑高速分析、反馈、判别、决断的过程。不只是常识，还涉及对外部信息的感知和意识。比非直觉准。

直觉过程

1. 选择性编码：从原始信息中 ***筛选*** 出与目标有关的信息
2. 选择性组合：将编码的信息以某种方式组合起来，形成具有合理性的内部联系的整体
3. 选择性比较：利用新的信息与记忆的信息的相似性更好的理解新信息

直觉推理的方法

启发信息：决定方向，来源于经验或内生

参考点：初始迭代，依赖于对其他相关事物的参考

直觉决策 ***不是*** 寻求目标绝对解的位置，而是评估某一参考点位置的选择是否更利于**损失的回避**

认知地图

过去经验中形成的一种对于局部环境的综合表象。一种认知映射。

直觉推理与认知映射的关系

认知地图（决策库）-搜索决策-决策与任务匹配-直觉反应

直觉的作用：对决策搜索的引导以及对代价空间的构造。

AlphaGo

离线学习

1. 棋谱→CNN决策网络&线性快速走棋网络
2. 互相博弈→增强的策略网络
3. SL 前U-1步，随机采样U，RL自我博弈直到结束。对第U步训练价值网络

SL: Supervised Learning

RL: Reinforcement Learning

在线对弈

核心思想：MCTS中嵌入DNN减少搜索空间

1. 提取特征
2. 策略网络估计各点走子概率
3. 由概率得权重（各边初始权重）
4. 利用价值网络和快速走子网络的自我对弈分别判断局势，获得各点得分（Rollout和价值估计）
5. 更新权重然后MCTS。某结点访问次数大于阈值，扩展MCT，重复上述步骤继续搜索。

直觉的体现

1. 策略网络：落子棋感
2. 价值网络：胜负棋感
3. MCTS：搜索验证

因果关系

一个事件（原因）导致了另一个事件发生（结果）

因果关系发现

推断因果关系

因果表征学习

机器学习方法学习因果关系，得到更加**准确**和**可解释**的模型。帮助人类理解。

神经图灵机（NTM）

Neural Turing Machine
特点：

1. 将外部内存资源与神经网络耦合。注意力形式交互。
2. 端到端可微。
3. 计算机程序三种基本机制：基础运算、逻辑流控制、计算过程中可读写的内存。现代机器学习只关注基础运算。RNN具有图灵完备性，可以模拟任意过程。但RNN模拟实现较困难，因此提出NTM。

与传统神经网络的不同

传统神经网络只通过输入输出与外界交互。

NTM还能通过选择性的读写操作与内存矩阵进行交互。具有简单的记忆与推理功能。

与传统图灵机的不同

传统图灵机的读写是针对内存中的某一具体位置。

NTM是模糊的读写。在读写时会根据权重与所有元素进行交互。

（符合深度学习模型的输出：带有一堆大小不一的权值的张量）

生成式对抗网络（GAN）

生成模型尝试欺骗判别模型

判别模型尽量保持不被欺骗

注意损失函数的记忆