复习提纲

1. **智能体BDI（belief desire intention）模型**

* 概念和逻辑上的理论模型，研究Agent理性和推理机制的基础
* 分为静态和动态两部分
  + 静态部分：信念、愿望（目标）和意图
  + 动态部分：BDI模型的推理过程

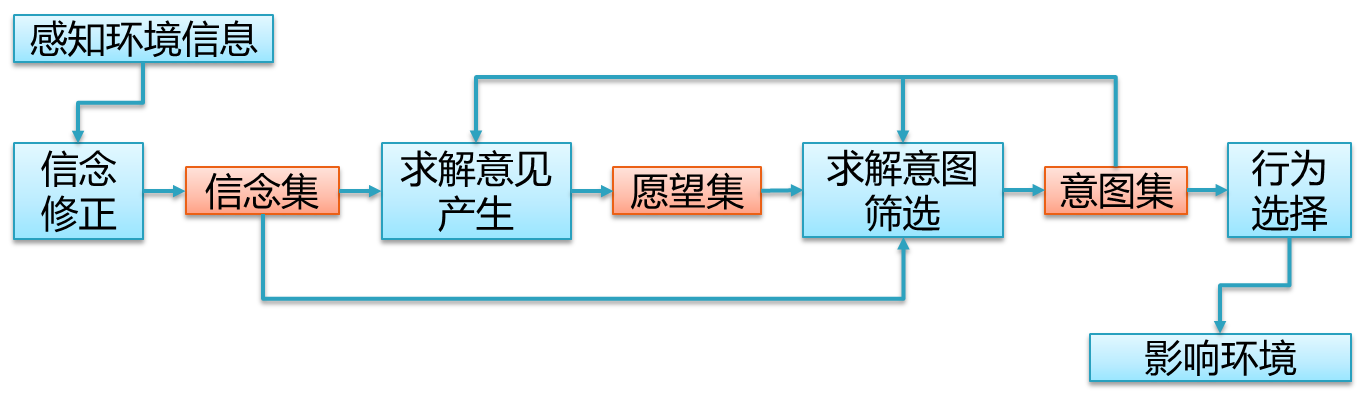


图 1 BDI Agent的一般结构

* 信念是关于外部世界以及自身状态的信息的一种评价，并随着时间的推移和环境的变化逐渐积累和更新。信念的内容不仅仅与外界环境的真实变化有关，还取决于实体对变化的理解，由于获取信息方式和能力不同，同样的外部环境下不同的 BDI模型实体的信念完整度、准确性不同。实体的所有的信念组成了信念集。
* 愿望是实体希望达到的事件状态，通常被理解为目标。实体的愿望往往不是只有一个，而是一个愿望集合，集合中的元素可以是在实现后不能共存的，因为愿望只是为实体的动作的产生提供了一个参考方向，并不会直接产生动作。与信念一样，愿望也是不断更新的，已经实现或者已经确定无法实现的愿望将被从愿望集中删除，新的愿望会不断产生并添加到愿望集中，但愿望总会持续一定时间，否则该愿望就不曾存在。
* 意图是实体决定为实施动作的事件状态，意图与动作的产生有直接的因果关系。与愿望不同的是，意图得持续时间小于愿望。意图也可以存在多个，组成一个意图集，并且意图集也是不断更新的，已经实现的意图会被从意图集中删除，新的意图会添加到意图集。

BDI模型在推理时，有四个动态转化过程：

* 信念修正：根据外界环境的变化信息更新当前的信念集
* 愿望产生：根据信念集的变化更新愿望集
* 意图生成：从愿望集中筛选出最适宜实现的且不冲突的愿望子集，制定实现愿望的规划，产生新的意图，并删除已经达到的意图。
* 行为选择：选择能实现意图的直接动作

1. **多智能体系统（Multi-Agent System）**

MAS是由多个Agent个体构成的系统。其中的Agent是一种具有一定自治能力的智能实体，它们通过相互作用，用以追求某些目标或完成某些任务，即由多个相互操作、相互作用的Agent构成的系统。这种以**“分而治之”“相互协作”**为基本特征的MAS使复杂系统问题的解决成为可能。

1. **多智能体系统的体系结构分类**

（1）集中式结构

至少拥有一个管理机构负责对组内的所有Agent成员的行为、协作、任务分配以及共享资源等提供统一的协调和管理服务。管理机构与各Agent成员之间具有一定程度的管理与被管理的关系

（2）分布式结构

Agent之间是平等关系，相互提供服务。各Agent的任务划分与佳酿、共享资源的分配和管理、冲突的协调、行为的一致性等，在遵循可能的社会规则和共享资源的管理策略基础上，由各Agent通过彼此的相互作用和对所处环境的感知，运用其自身的知识进行合理的判断和推理自己作出决策来实现。

（3）混合式结构

由集中式和分布式两类结构混合组成。包含一个或多个管理服务机构，此机构只对部分成员Agnet以某种方式进行统一管理，其他成员Agent之间是平等的，它们所有行为由自身作出决策。适应分布式MAS复杂、开放的特性，目前普遍采用。

1. **多智能体的通信方式**

直接发送、广播、黑板结构和公告板。

1. **多智能体系统冲突形式**

任务冲突、空间冲突、信息冲突

1. **合同网协议**

用于解决分布式问题求解环境下各节点之间任务分配而进行的一种合约协作过程。两个节点之间就任务的委托和承揽构成合同关系，一组这样的节点就构成合同网。

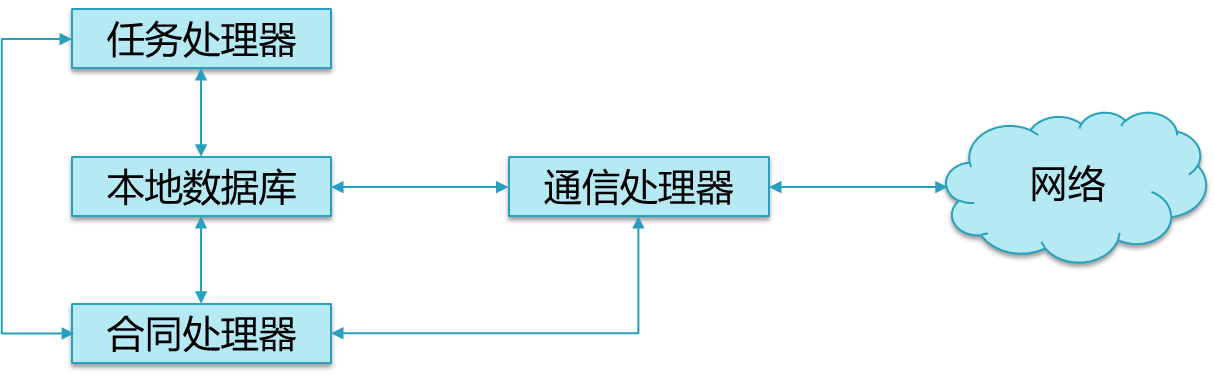


图 2 合同网协议的一般结构

本地数据库包括与节点有关的知识库、协作当前状态和问题求解过程的信息。另外三个部件利用本地知识库执行它们的任务。通信处理器与其它结点进行通信。结点仅仅通过该部件直接与网络相接。特别是通信处理器应该理解消息的发送和接收。合同处理器判断发标所提供的任务、发送投标和完成合同。它也分析和解释到达的消息。最后，合同处理器执行全部结点的协作。任务处理器的任务是实际处理赋予它的任务的求解。它从合同处理器接受所要求解的任务利用本地数据库进行求解并将结果送到合同处理器。

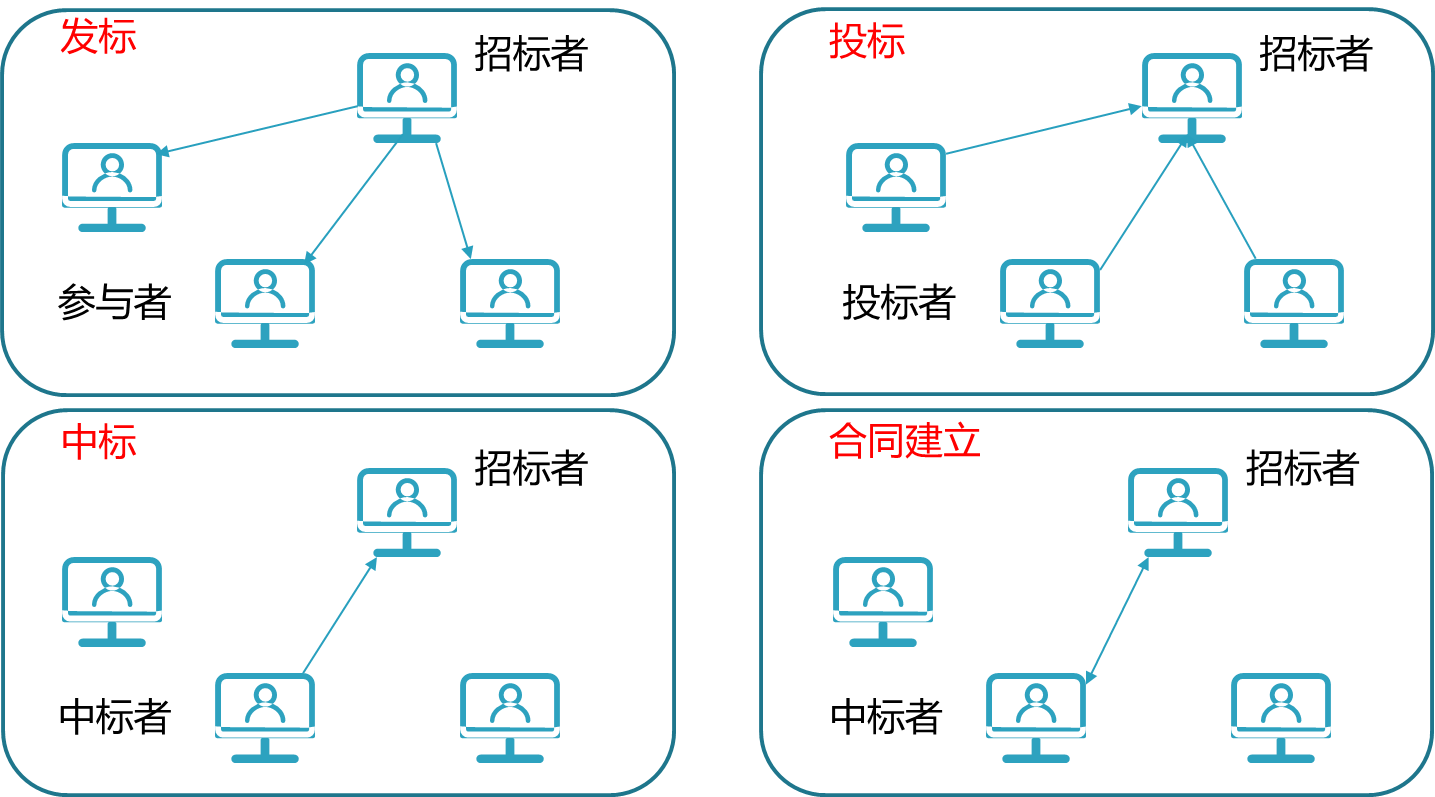


图 5 合同网协议的工作过程

1. **信息融合定义**

一种多层次、多方面的信息处理过程，包括对多源数据进行检测（detection）、相关（correlation）、组合（combination）和估计（estimation），从而提高位置估计（position estimation）和身份估计（identity estimation）的精度，以及对环境态势评估（situation assessment）和威胁估计（threat assessment）及其重要程度进行适时的完整评价。

要点：

* 多传感器系统是信息融合的硬件基础
* 多源信息是信息融合的加工对象
* 协调优化和综合处理是信息融合的核心

1. **信息融合结构**

检测级结构模型包括并行模型、串行模型、树状模型等

位置级结构模型包括集中式、分布式、混合式等

属性级结构模型应用较为广泛，它根据整合过程在系统中位置的不同，将信息融合过程划分为数据层融合、特征层融合和决策层融合三个层次

1. **数据层融合**

直接对观测数据进行融合，是最低层次的融合

不存在信息丢失问题，能够提供较多的细节信息

一般要求多源数据来自于同质传感器，并且对系统的数据处理能力和通信带宽要求很高

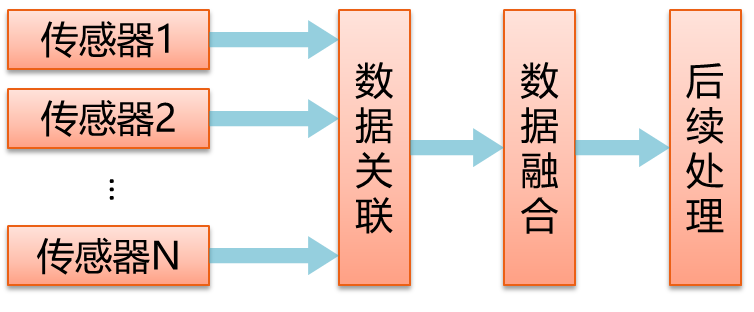


图 4数据层融合模型

1. **特征层融合**

首先从观测数据中提取代表性特征，然后对特征进行融合

属于中间层次的融合

对系统的数据处理和通信能力的要求相对较低

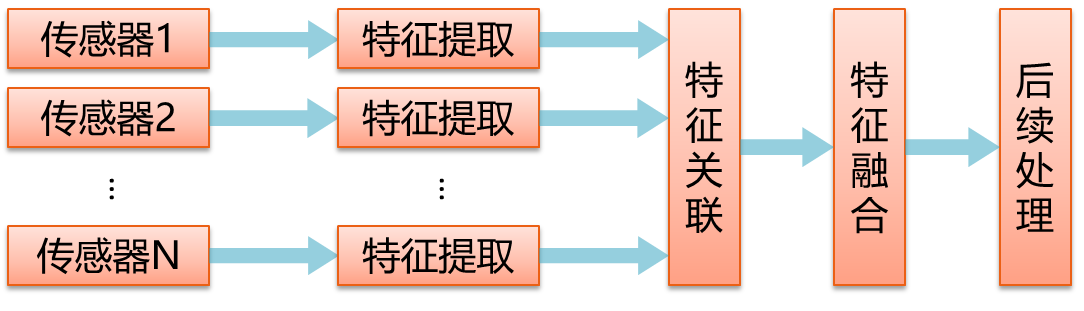


图 5 特征层融合模型

1. **决策层融合**

每个数据源首先对各自的数据作出决策，然后将多个决策结果进行融合

属于高层次的融合

具有通信量小、抗干扰能力强、对传感器依赖程度低等特点

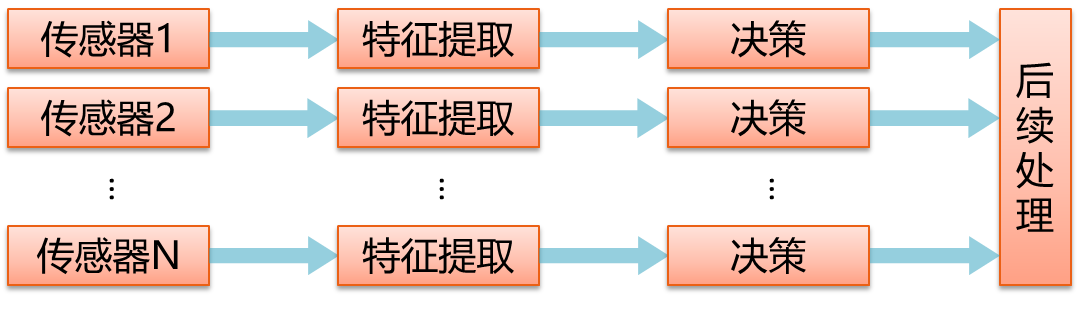


图 6 决策层融合

1. **OODA模型**

包含观测（observe）、定向（orient）、决策（decide）和行为（action）四个模块。

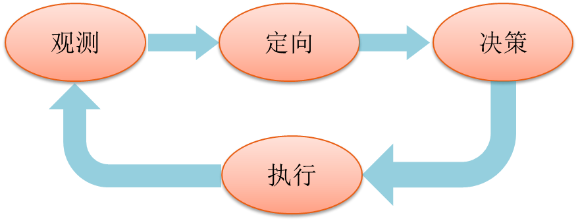


图 7 OODA模型结构图

1. **Agent体系结构**

（1）慎思型（deliberative）Agent

将Agent看作一种特殊的知识系统，即通过符号AI的方法来实现Agent的表示和推理，侧重研究Agent信念、愿望、意图关系以及它们的形式化描述。慎思型Agent具有较高的智能，但反应慢、执行效率低。

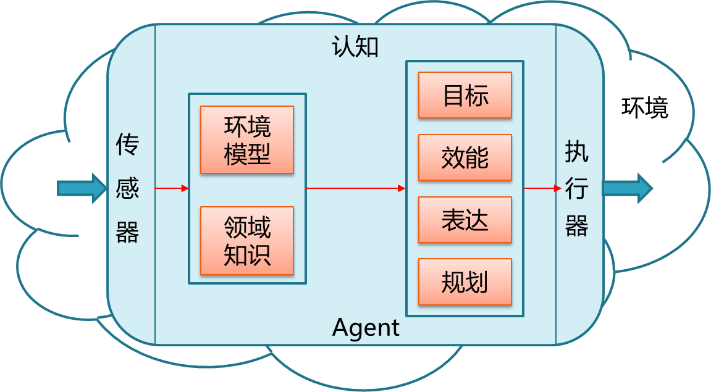


图 8 慎思型Agent

（2）反应型（reactive）Agent

Agent的智能取决于感知和行动，从而形成“感知-动作”模型。结构简单，但实践中非常高效。反应型Agent虽然反应快、能迅速适应环境变化，但智能水平较低。

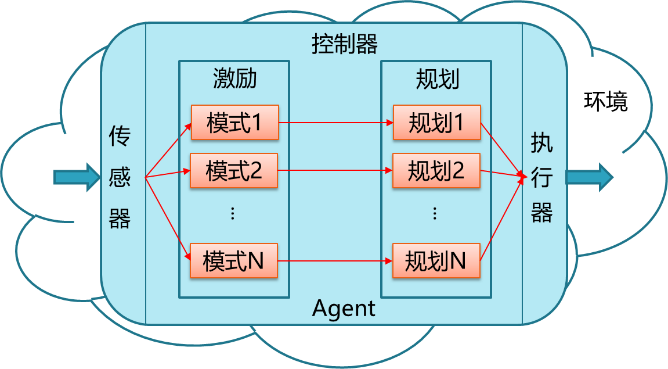


图 9 反应型Agent

（3）混合型（reactive）Agent

结合了慎思型和反应型两种Agent的特点，通常包括两层结构：

* 高层：认知层，用传统符号AI的方式处理规划和进行决策
* 低层：反应层，能快速响应和处理环境中突发事件，不使用任何符号表示和推理系统，通常优先级更高

1. **随机博弈（Markov博弈）**

在单智能体系统中智能体的动态决策其实是一个Markov决策过程，而在多智能体系统中智能体的Markov决策过程的扩展形式就是随机博弈（Markov博弈）。

每个智能体面临的是一个不同的Markov决策过程，这些智能体的Markov决策过程通过它们的奖励函数以及依赖于智能体联合行动的系统动态特性交织在一起。

Markov博弈以Nash平衡作为协作的目标。

是多个玩家参与，有多个状态的博弈，具有Markov特性。可以表示为一个元组。

其中：

* 为转移函数，指智能体当前状态和联合行为时下一状态的概率分布
* 为折扣因子
* 是智能体𝑖的回报函数，表示智能体𝑖在状态𝑠采取联合行为之后在状态所得的回报

1. **极大极小Q学习算法**

极大极小Q学习算法采用极大极小原理来求解玩家的纳什均衡策略以及双人零和随机博弈的状态值。与Q学习算法类似，极大极小Q学习算法是一种对状态值或状态—行为值执行反向传播的时间差分学习方法。极大极小Q学习算法表述如下：

在双人零和随机博弈中，给定当前状态、定义玩家的状态—值函数为

式中，表示玩家的对手；表示在状态下玩家的所有可能策略；为玩家及其对手分别选择行为和且之后采用纳什均衡策略时的预期回报。

极大极小Q学习算法如算法1所示，其中每次迭代时必须采用线性规划来求解和。

|  |
| --- |
| **算法1** **极大极小Q学习算法**  1. 初始化、和  2. 对于每次迭代  3. 玩家根据探索—开发策略从当前状态采取行为  4. 在随后的状态中，玩家观测获得的回报以及对手在前一状态所采取的行为  5. 更新：  式中，为学习速率；为折扣因素  6. 利用线性规划求解  并得到更新后的和  7. 结束 |

可以用图形解法求解步骤6中的线性规划问题。

1. **Nash Q**

纳什Q学习算法，是将极大极小Q学习算法从零和随机博弈扩展到一般和随机博弈。在纳什Q学习算法中，需计算每个状态下的纳什Q值，以更新行为—值函数并得到均衡策略。尽管纳什Q学习是应用于一般和随机博弈，但收敛于纳什均衡的条件并不包括相应的一般类环境。相应的环境类别实际上仅限于只学习到协调均衡或对抗均衡的博弈情况。纳什Q学习算法如算法2所示。

|  |
| --- |
| **算法3.2** **纳什Q学习算法**  1. 初始化  2. 对于每次迭代  3. 根据探索—开发策略，玩家从当前状态采取行为  4. 在下一状态，玩家观测所有玩家所得到的回报，以及在先前状态所有玩家采取的行为  5. 更新：  (3.15)  6. 采用二次规划更新纳什和  7. 结束 |

1. **CTDE（centralized training decentralized execution）集中训练分散执行**

训练时使用中央控制器，辅助智能体做训练，训练结束后，不再需要中央控制器，每个智能体独立根据本地观测做决策。这是目前最流行的架构，有利于学到更好的策略，而去中心化决策无须通信，可以做到实时决策。

1. **基于值分解的多智能体强化学习方法**

是团队合作的多智能体强化学习的常用方法之一。主要是通过某种机制，将团队的Q分解到个体上，指导个体策略学习。VDN(Value Decomposition Networks)、QMIX、QTRAN等是这类方法的典型代表。重点是要解决以下几个方面的问题：

* “credit assignment”（信用分配）问题，即利用团队奖励拟合出的值函数不能评价每个智能体的策略对整体的贡献。
* “spurious reward”（虚假奖励）由于每个智能体都是局部观测，对其中一个智能体来说，其获得团队奖励有可能是其队友的行为导致的。所以，每个智能体独立使用强化学习算法往往效果很差。
* “lazy agent”（懒惰智能体）当团队中部分智能体学习到了比较好的策略并且能够完成任务时，其他智能体不需要做什么也能够获得不错的团队奖励。

1. **混合增强智能的定义**

混合增强智能包含两种形式：人在回路的混合增强智能与认知计算的混合增强智能。

（1）Human-in-the-loop (HITL) hybrid-augmented intelligence is defined as a intelligent model that requires human interaction. In this type of intelligent system human is always part of the system and consequently influences the outcome in such a way that humans give the final judgment if a low confident result is given by computer. Human-in-the-loop hybrid-augmented intelligence also readily allows for the treatment of problems and requirements that may not be easily training and classifying by machine learning.

人在回路（Human-in-the-loop, HITL）的混合增强智能是指需要与人进行交互的智能模型系统。在这种类型的智能系统中，人始终是系统的一部分，并以某种方式影响结果，如果计算机给出的结果置信度较低，则人可以做出最终判断。人在回路的混合增强智能还可以轻松的从场景中提取出问题和需求，而这些问题和需求可能无法通过机器学习轻松地进行训练和分类。

（2）At present, there is several different definitions for cognitive computing (CC) based intelligence in either academia or industry. In general, cognitive computing based hybrid-augmented intelligence has been used to refer to new software and/or hardware that mimics the functioning of the human brain and improves computer's capabilities of reasoning，decision-making and perception In this sense, the cognitive computing based hybrid-augmented intelligence is a new framework of computing with the goal of more accurate models of how the human brain/mind senses, reasons, and responds to stimulus，especially how to build causal model，intuitive reasoning model and associative memory.

当前，学术界或相关领域中基于认知计算（Cognitive Computing, CC）的智能有几种不同的定义。一般情况，基于认知计算的混合增强智能被用于指代模仿人脑功能并提高计算机推理，决策和感知能力的新软件和（或）硬件。混合增强智能是一种新的计算框架，其目标是建立更精确的模型，以建立关于大脑如何感知、推理和响应刺激的模型，尤其是如何建立因果模型、直觉推理模型和联想记忆模型。

1. **人在回路的智能场景定义**

把人的作用引入到智能系统的回路中，可以把人对模糊、不确定问题分析与响应的高级认知机制与机器智能系统紧密耦合（如下图所示），使得两者相互适应，协同工作，形成双向的信息交流与控制。把人的感知、认知能力和计算机强大的运算和存储能力相结合，可以形成“1+1>2”的增强智能形态，从而实现大规模的非完整、非结构化知识信息的处理，同时避免由于当前人工智能技术的局限性而带来的决策风险和系统的失控等问题。

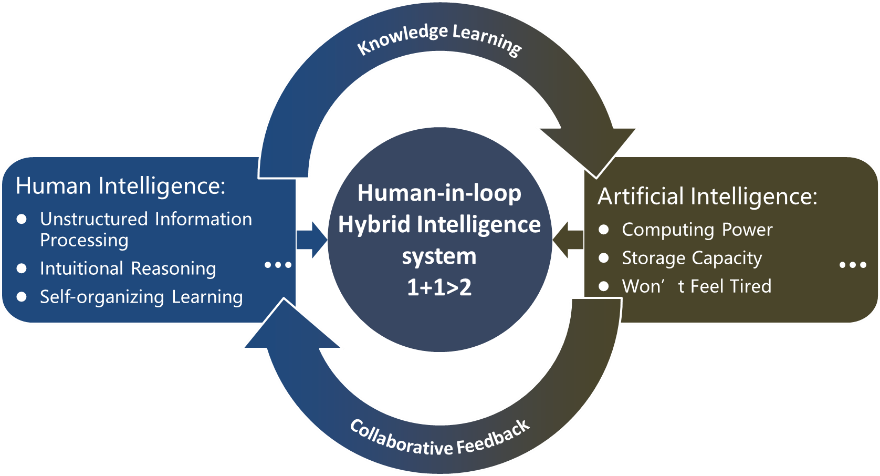


图 10 人在回路的智能场景示意图

1. **人在回路的混合-增强智能的一种实现框架**

不同领域对应不同的人在回路的混合-增强智能的构建。下图给出了人在回路的混-增强合智能的一种基本框架，该框架可以看作是一种混合学习模式，它集成了机器学习、知识库和人类决策，它利用机器学习（有监督、无监督的）从训练数据或少量样本中学习得到模型，并将该模型应用于新的数据给出预测。当预测的置信度不高时，人类将介入做出判断。在下图所示的混合学习框架中，当系统出现非典型或异常的情况，或计算机缺乏成功把握时，由置信度评估或由计算机的认知负载的状态确定是否由人类调整预测，或由人类主动介入，同时自动更新系统的知识库。由于人的预测介入到算法中，从而使系统提高了准确度和可信度。人在回路的混合智能应尽可能减少人类在系统中的参与度，使计算机能完成大部分工作。下图给出的混合学习模式的智能可以大大扩展人类能够完成任务的尺度和任务。

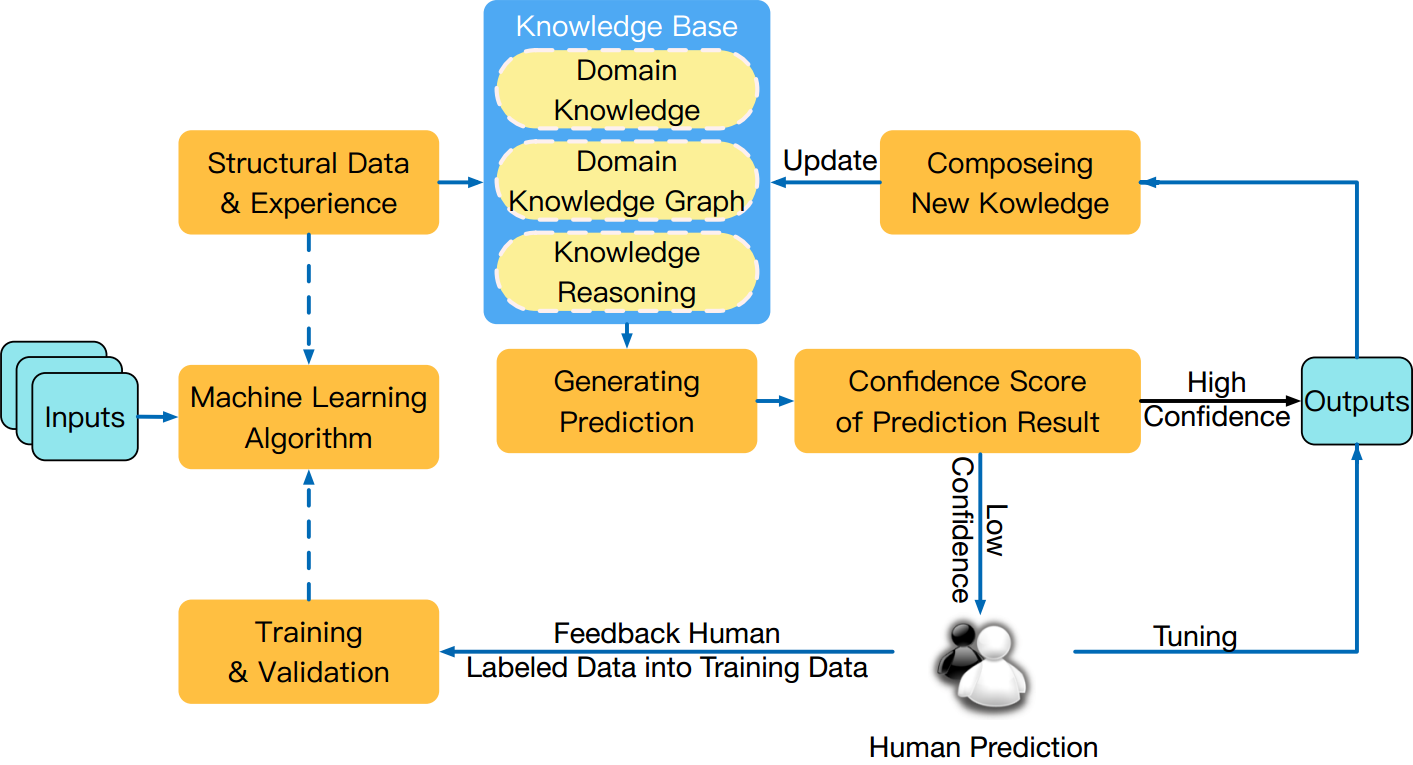


图 11 人在回路的混合-增强智能的一种实现框架

1. **实现人在回路的混合-增强智能的主要研究内容**

（1）如何使人在智能回路中用自然的方式训练机器，突破人机知识交互屏障；

（2）如何将人类的决策、经验与智能机器在逻辑推理、演绎归纳等计算处理的优势相结合，实现高效的、有意义的人机协同构建；

（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推动的机器学习方法，使机器不仅能从海量训练样本学习，还能从人类知识中学习，并利用学习的知识完成高度智能化的任务。

1. **认知计算框架**

认知计算框架的基本组成包含感知、注意、理解、证实、规划和评测6个相互关联的认知环节（如下图所示），其中任何一个都可视为一个认知任务的起点或目标。根据满足目标任务所需要信息与外界不断交互，从而选择简单路径还是复杂路径（如多次迭代），达到认知任务的目标。通常，自上而下的选择性注意是基于规划的注意，而自下而上的选择性注意本质上是基于感知的注意。基于理解或者规划下的评测为先验概率（表现为预测），基于感知下的评测为后验概率（表现为实测）。简言之，认知计算的过程就是根据满足目标任务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（计算），而非孤立的限制于“基于确定知识的处理”。面对预先没有准备的问题，智能应当具有“做…直到”的循环能力，而不应该通过遍历每一种可能性来获得目标规划，这就要求认知计算的过程还应当自身蕴含“证实”的步骤：下一步该做什么？是否产生了预期的结果？以及是否有足够的进展表明应当在既有的策略下进一步努力，或者尝试其它方法；在此过程中，通过推理和经验（长时记忆）对环境产生丰富的理解和引导，并且形成以此为依据的“证实”能力。

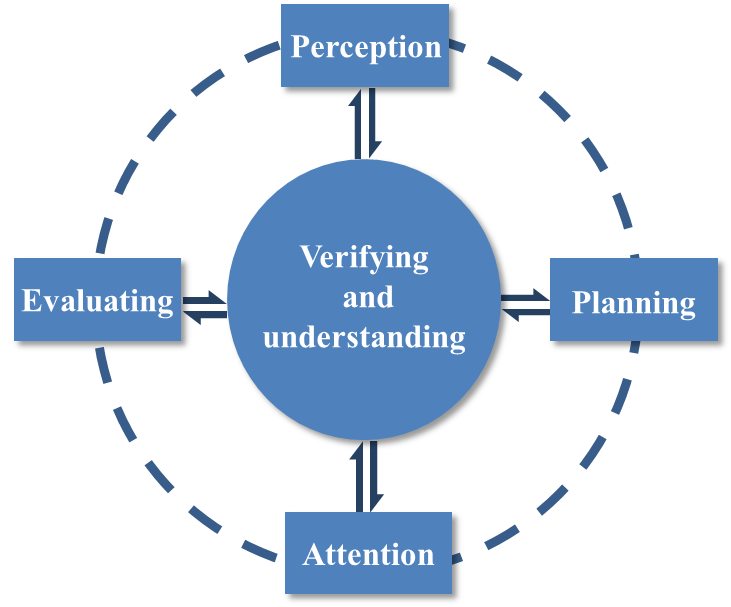


图 12 认知计算框架

1. **直觉过程**

直觉过程是人脑高速分析、反馈、判别、决断的过程。人类许多时候都在进行综合风险判断，但这常常是不耗费精力的。研究表明，人类的直觉判断平均正确率比非直觉判断更高。人类习惯于在日常生活中做出直觉决策，如：判断两个物体的远近，察觉对方语气中的不友善，选择生活伴侣，选择一本要读的书，等等。直觉决策不是只使用常识，它还涉及到对来自外部信息的感知和意识。

1. **直觉过程的划分**

直觉可以分成几个相对独立又相互联系的过程，即选择性编码、选择性组合与选择性比较。选择性编码是从隐含在大量与目标无关的原始信息中筛选出与目标有关的信息。仅由选择性编码不足以产生正确的认知，还需要有选择地把已编码的信息以某种方式组合起来，形成一个具有某种合理性的内部联系的整体，这就是直觉中的选择性组合。选择性比较是利用新的信息与记忆的信息在某种程度上的相似性，去更好地理解新信息。

1. **直觉推理的方法**

直觉推理依赖的是启发式、参考点方法。启发信息（先验信息）决定着问题求解的方向，它来源于经验或内生；参考点决定问题求解的 初始迭代，它的确定依赖于对其他相关事物的参考。因此，直觉决策不是寻求目 标求解的绝对解位置，而是评估某一参考点位置的选择是否更有利于损失的回避。

1. **认知地图**

认知地图（又称心理地图或心智模式）是一种在过去经验中形成的，一种对于局部环境的综合表象，即包含事件的简单顺序，也包括方向、距离，甚至时间关系。从信息加工理论的角度来看，认知地图实质上是一种认知映射（cognitive mapping），它是一个包括获取、编码、存储、内部操作、解码和使用外部信息的动态过程。

1. **直觉推理与认知映射的关系**

人的认知活动体现在基于“认知映射”的一系列决策活动中，这是一个不断地模式匹配的过程。当前认知映射的形成与大脑对外部信息的感知和理解相关。下面用一种简单回路的方式来说明直觉与认知映射的关系。如下图所示，人类个体在成长过程中通过学习、常识和经验的积累形成了“决策库”，人脑随机地在“决策库”中搜索决策，一旦被选中的决策与当前认知映射过程中的任务匹配（匹配的度量可以是最小代价回避损失），人就做出了直觉反应。在这个过程中，直觉的作用可以看作是在计算过程中对决策搜索的引导以及代价空间的构造。

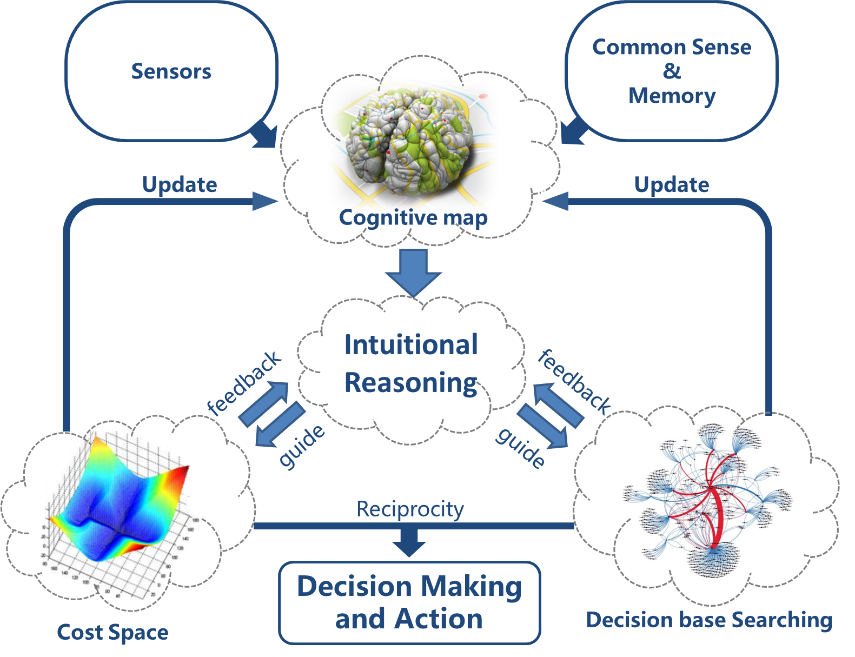


图 14 直觉推理与认知映射的关系

1. **AlphaGo算法的基本组成**

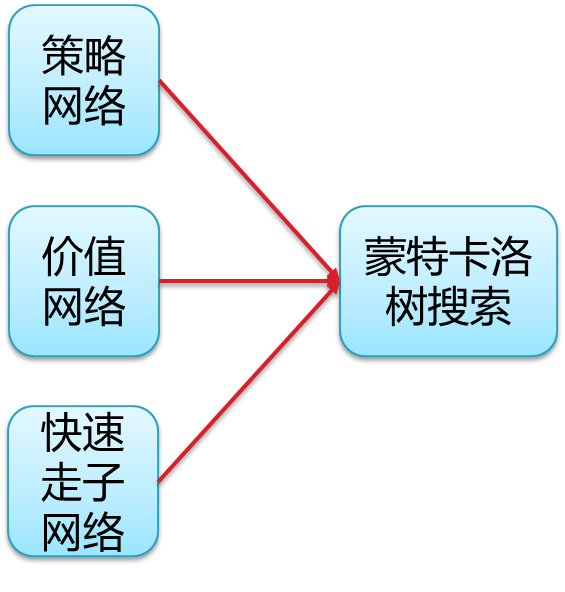


图 15 AlphaGo的基本组成

1. **AlphaGo的训练**

离线学习（1）

* 利用3万多幅专业棋手对局的棋谱来训练两个网络。
* 一个是基于全局特征和深度卷积网络（CNN）训练出来的策略网络（Policy Network）。其主要作用是给定当前盘面状态作为输入，输出下一步棋在棋盘其它空地上的落子概率。
* 另一个是利用局部特征和线性模型训练出来的快速走棋策略（Rollout Policy)。
* 策略网络速度较慢，但精度较高；快速走棋策略反之。

离线学习（2）

* 利用第t轮的策略网络与先前训练好的策略网络互相对弈，利用增强式学习来修正第t轮的策略网络的参数，最终得到增强的策略网络。

离线学习（3）

* 先利用普通的策略网络来生成棋局的前U-1步，然后利用随机采样来决定第U步的位置。
* 利用增强的策略网络来完成后面的自我对弈过程，直至棋局结束分出胜负。
* 第U步的盘面作为特征输入，胜负作为label，学习一个价值网络（Value Network），用于判断结果的输赢概率。

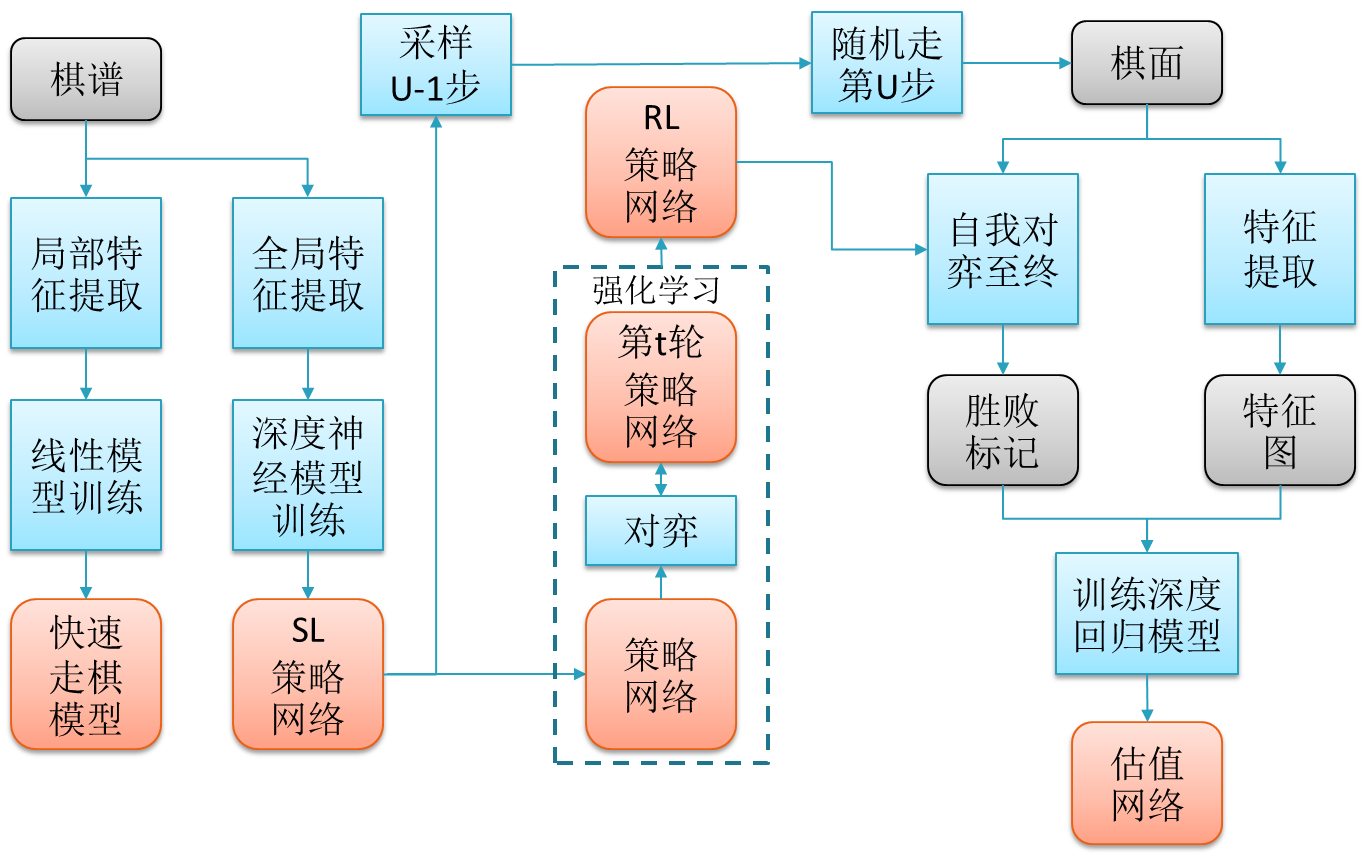


图 16 AlphaGo的训练过程

1. **AlphaGo的在线对弈**

其核心思想实在MCTS中嵌入了深度神经网络来减少搜索空间。

（1）根据当前盘面已经落子的情况提取相应特征；

（2）利用策略网络估计出棋盘其他空地的落子概率；

（3）根据落子概率来计算此处往下发展的权重，初始值为落子概率本身。

（4）利用价值网络和快速走棋网络分别判断局势，两个局势得分相加为此处最后走棋获胜的得分。

（5）利用第四步计算的得分来更新之前那个走棋位置的权重；此后，从权重最大的0.15那条边开始继续搜索和更新。这些权重的更新过程应该是可以并行的。当某个节点的被访问次数超过了一定的门限值，则在蒙特卡罗树上进一步展开下一级别的搜索。

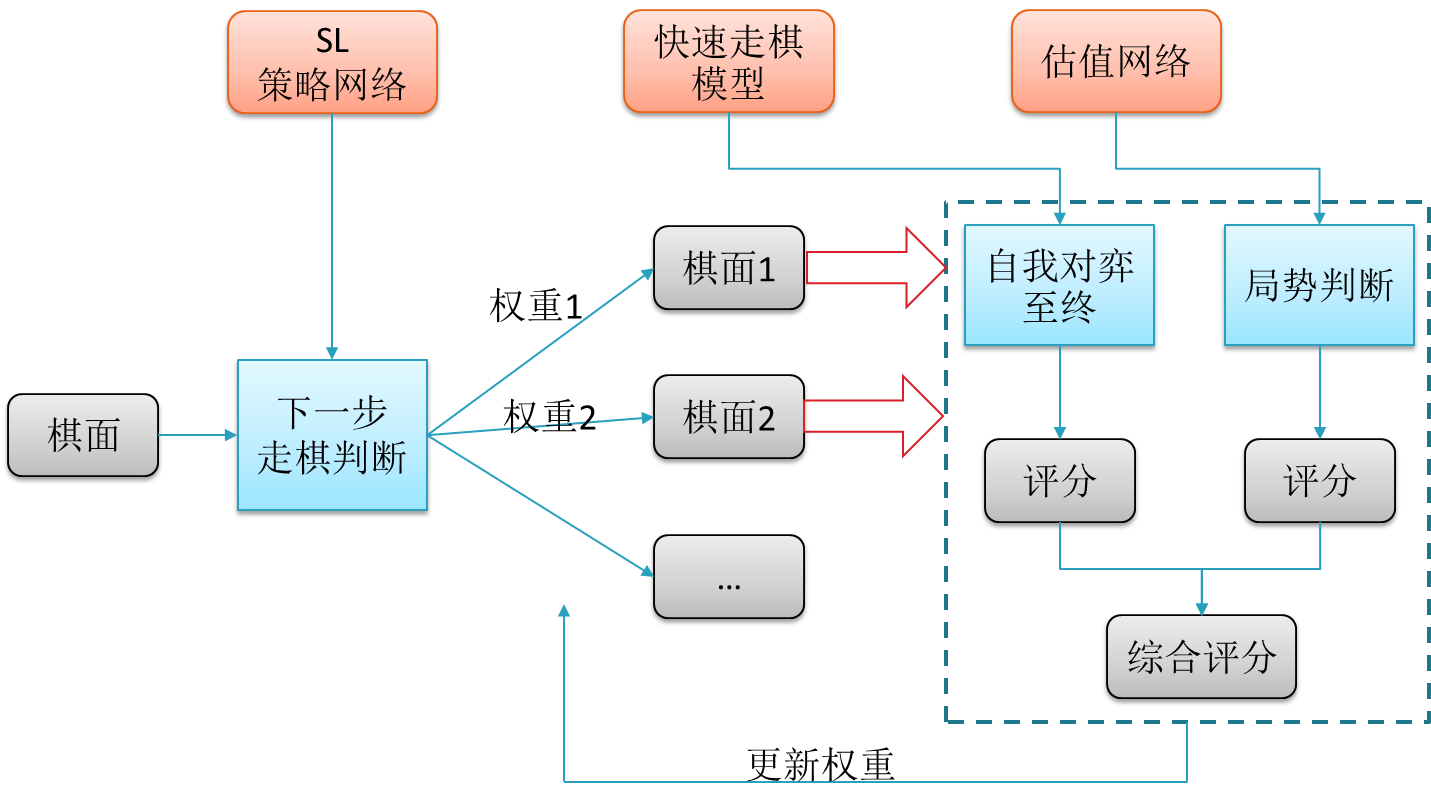


图 17 AlphaGo的在线对弈

1. **直觉在AlphaGo中的体现**

（1）策略网络：落子棋感

* 有监督学习， 即直接对已有的棋手对弈棋谱进行不断的学习, 从而获得棋手的棋感.。
* 策略网络学习的对象是利用围棋知识从当前盘面中提取的多个特征矩阵，每个盘面都有一个标示, 这个标示标记了该盘面下一步棋手是如何落子的。
* 采用卷积神经网络，结构相对简单、训练参数少和适应性强。
* 训练集由职业棋手和业余高端棋手的十几万份棋谱组成，包含上千万和落子方式。
* 策略网络预测的正确率可以达到57%,，需要在准确度和速度上取得平衡。
* 快速走子模块与策略网络基本相似，输入特征少，并采用简单的线性结构，计算速度快。准确率为24.2%。

（2）价值网络：胜负棋感

* 数据集是利用强化后的策略网络进行三千万盘的自我博弈形成的棋谱， 通过学习三千万盘棋的胜负结果进行学习和更新，从而获取在围棋盘面的胜负棋感。
* 价值网络的输入与策略网络基本相同，不过在策略网络的基础上又添加了现在要走子的颜色这一特征矩阵。

（3）蒙特卡罗树搜索：搜索验证

* 对落子棋感和胜负棋感进行了计算验证， 最终得到最为有效的结果。
* 对节点进行动态评估，根据评估结果指引对搜索树的选择
* 评估结合了策略网络和价值网络的结果

1. **因果关系**

因果关系是指一个事件（原因）导致了另一个事件（结果）发生的关系。在因果关系中，原因是导致结果发生的必要条件。因此，如果我们改变原因，我们也会改变结果。因果关系是科学研究和决策制定中非常重要的概念，因为它可以帮助我们理解事件之间的关系，并预测可能的结果。

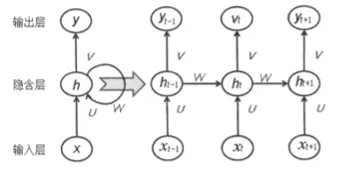
1. **因果关系发现**

因果关系发现（causal discovery）是指从观测数据中推断出变量之间的因果关系的过程。在纷繁复杂的数据中，很难直接观察到变量之间的因果关系，因此需要通过特定的算法和方法来进行推断和分析。因果关系发现是现代数据科学领域中的一个重要研究方向，可以应用于医学、社会科学、金融等领域。

1. **因果表征学习**

因果表征学习是一种机器学习方法，旨在从数据中学习变量之间的因果关系，以获得更加准确和可解释的模型。与传统的机器学习方法不同，因果表征学习强调变量之间的因果关系，而不是仅仅寻找它们之间的相关性。通过学习因果关系，我们可以更好地理解数据背后的机制，并且能够更好地预测未来的结果。因果表征学习在医疗、金融、社会科学等领域都有广泛的应用。

1. **循环神经网络（RNN）**

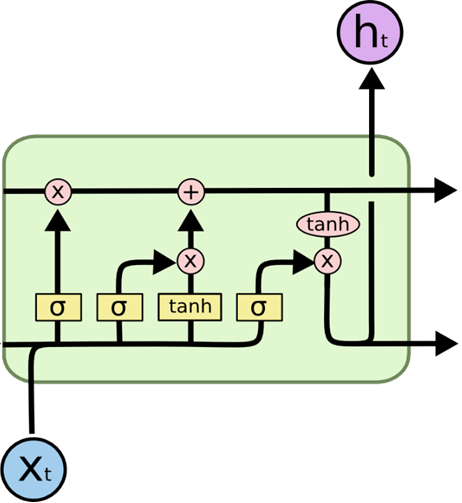


= t时刻的输入

= t时刻的输出

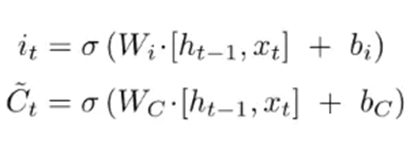
= t时刻的隐状态

1. **长短时记忆网络（LSTM）**



1. 输入门

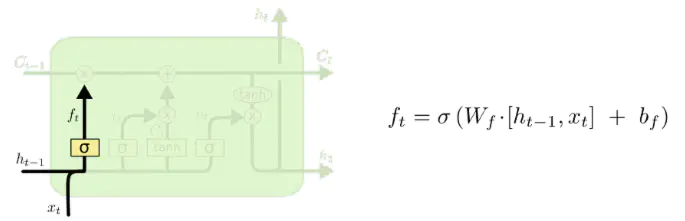
输入门是决定我们要在细胞状态中存储什么信息，它分为两部分，第一部分是决定哪些信息要输入，其次是要输入的具体信息，然后组合来创建更新值。输入门也是由前一个状态的隐藏值也就是ht-1和当前时刻的输入xt来共同决定的



1. 遗忘门

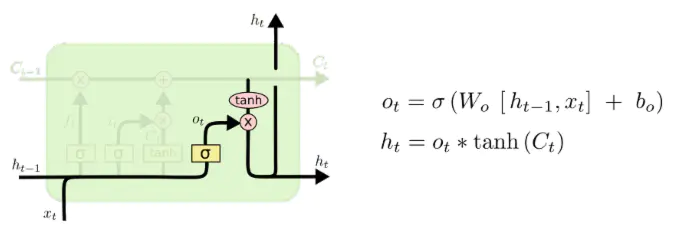
遗忘门决定我们要从细胞状态中丢弃什么信息，他是由前一个隐藏状态ht-1和当前输入xt来共同决定的，他为上一个状态ct-1中的每个数字输出0和1之间的值，表示我们要从前一个状态中丢弃什么信息，以一个句子为例，“细胞状态可能包括当前主语的性别，从而决定使用正确的代词(他/她/它)，当看到新的主语时，就需要忘记旧主语的性别。”也就是从我们当前的输入来决定应该忘记哪些历史信息。

当前时刻的细胞状态Ct等于遗忘门乘上一时刻状态加上输入门乘当前输入的信息就得到了最新的状态。

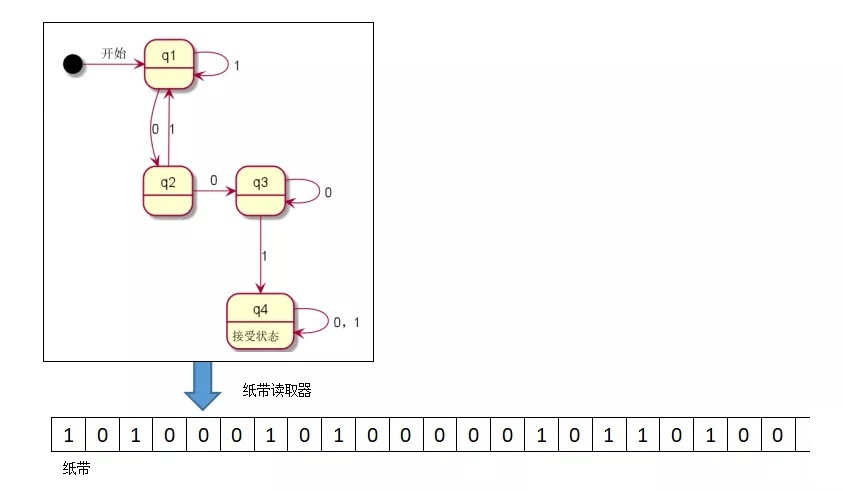


1. 输出门

输出门，就是要决定输出哪一部分信息给下一时刻，首先有一个门控， 输出0-1之间的值，然后将细胞状态通过tanh操作乘以这个输出门得到结果，就完成了输出门决定将哪些信息给下一时刻。



1. **图灵机**



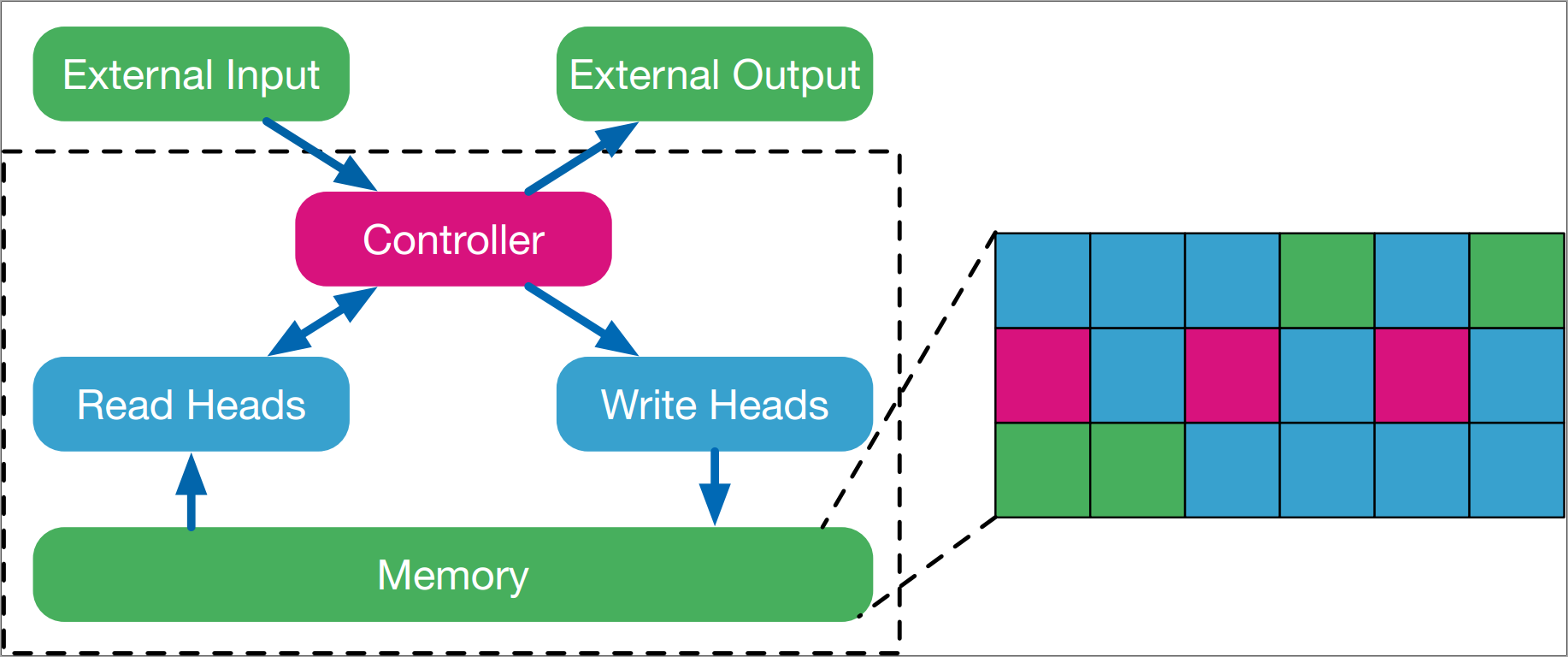
信息存储在纸带上

纸带读取器可以逐格读取纸带上的内容，每次一格

状态寄存器：有存储功能的零件，假设只有一个格子，存储一个状态

指令：当寄存器的状态是，且读到的纸带上的字符是，那么把状态寄存器的状态修改为

1. **神经图灵机**



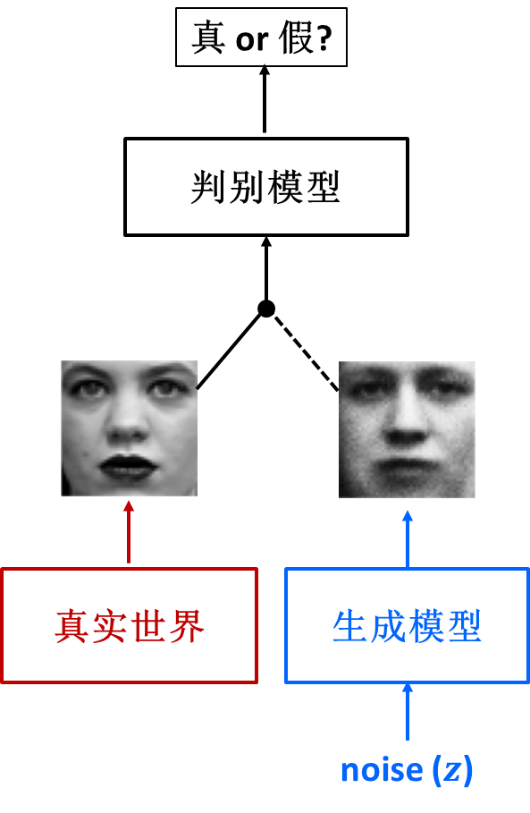
原文通过将外部内存资源(External Memory Resources)与神经网络耦合，扩展了神经网络的功能，神经网络将以注意力的形式与外部内存资源进行交互。结合后的系统类似于图灵机或冯诺依曼体系结构，只不过整个系统是端到端可微的(End-to-end Differentiable)，因此可以有效地使用梯度下降的方法进行训练。

计算机程序一般都会利用三种基本的机制：基础运算（如算术运算）、逻辑流控制（如分支）和计算过程中可读写的内存。现代的机器学习通常忽略了后两者。

RNN可以学习并执行对时间序列数据的复杂转换，此外RNN还具有图灵完备性，因此，在合理的设定下，RNN可以对任意一个过程(procedure)进行模拟（如模拟一个算法的执行）。然而直接利用RNN来解决算法执行的问题实现起来比较困难，因此原文中提出了一种新的结构——神经图灵机(NTM)，来增强传统RNN的能力，从而简化算法任务的学习和执行。

* 与传统神经网络的不同
  + 传统的神经网络只通过输入和输出向量与外部世界进行交互
  + NTM还能通过选择性的读写操作与内存矩阵进行交互。神经图灵机同时会与一个带有选择性读写操作的内存矩阵进行交互。因此，神经图灵机可以实现简单的记忆与推理功能。
* 与传统图灵机的不同
  + 传统图灵机的读写操作(即计算机中的读写操作)都是针对内存中具体某个位置的元素进行读写
  + NTM进行的是模糊(blurry)的读写，在读写的过程中会与内存中所有的元素进行交互，只不过对于每个元素的操作都有一定的权重
  + 例如在写入操作时，有的元素可能被很大程度的修改，而有的元素只被轻微地修改，由注意力机制来确定权重

1. **生成式对抗网络**



* 生成模型尝试欺骗判别模型
* 判别模型尽量保持不被欺骗
* 两个模型同时训练
  + 当生成模型变好时，判别模型判别难度增加
  + 当判别模型变好时，生成模型要求变高
* 最终，不关心判别模型

是要通过训练使得生成模型更好

损失函数：