

Artificial Intelligence: A Modern Approach

搜索

- 无信息搜索包括一致代价搜索、广度优先搜索、深度优先搜索
- 除了问题本身的定义外，使用问题特定知识的搜索策略被认为是有信息搜索、启发式搜索
- 某些 NP 完或 NP 难问题只能通过搜索求解
- 问题形式化是给定一个目标，决定要考虑的动作与状态的处理
- 内存中仅保存一个节点似乎是对内存限制问题的极端反应，局部束搜索保持 k 个状态而不仅仅为一
- 遗传算法是随机束搜索的一个变体，其中后继节点的生成是由组合两个双亲状态，而不是修改单一状态
- 爬山搜索有时也被称作贪婪局部搜索，在三种情况下经常被困：局部最大值、山岭、高原
- 以下关于模拟退火算法陈述正确的是：
 - 模拟退火算法不是选择最佳行动，而是选择随机行动
 - 模拟退火算法的内循环与爬山法相似
- 蚁群优化算法是受蚂蚁在蚁巢和食物源之间寻找路径行为的启发而形成的
- 粒子群优化算法采用若干粒子构成一个围绕搜索空间移动的群体来寻找最优解
- 局部搜索算法使用一个当前节点（而不是多条路径）并且通常仅移动到该节点相邻的节点
- 除了寻找目标之外，局部搜索算法对解决纯优化问题也很有效。其目的是根据一个目标函数找到其最好的状态
- 对抗搜索通常称为博弈

- 非零和博弈中智能体交互动作的总收益和损失可以小于或大于零
- 下列关于零和博弈 maximum 概念中选择正确答案：
 - 每个玩家会是自己可能的最大收益变得最大
 - 每个玩家会使对手可能的最大损失变的最大
- Alpha-beta 剪枝旨在减少其搜索树中由 minimax 算法评价的节点数量
- Alpha-beta 剪枝旨在消除其搜索树中由 minimax 算法评价的大部分
- 程序应该早一些剪断搜索，并在搜索中对状态应用启发式评价函数，有效地将非终端节点 转换为终端叶节点
- 随机博弈是一种具有概率转换的动态博弈，有一个或多个玩家
- 蒙特卡罗方法是一大类计算算法，它凭借重复随机采样来获得数值结果
- 蒙特卡罗树搜索对最有利的动作进行分析，根据搜索空间的随机采样来扩展搜索树
- 用于约束满足问题 CSP 使用因子表示状态
- 设{A, B, C, D}为变量，每个变量的域是{u, v, w}，且 "!=" 表示不等于，从如下表达式中选择哪个是 CSP 形式化的 2 元约束？ Diff(A, D)
- CSP 与状态空间搜索的比较：
 - CSP 求解系统会比状态空间搜索求解系统快
 - CSP 可以快速排除大的搜索空间样本
- 哪些是约束传播局部一致性：弧一致、节点一致、路径一致、k 一致
- 如下有关回溯搜索正确的是：
 - 递增地构建解的候选，并且一旦确定部分候选 c 不能成为合法的解，就将 c 抛弃
 - 每次为变量选择值并且当变量没有合法赋值时回溯
- 约束满足问题 (CSP) 被定义为其状态必须满足若干约束和限制的一组对象

- 在对一个变量选择一个新值时，最小冲突启发式选择导致与其它变量呈现最少冲突的值
- 为了将约束图简化为树结构，有两种方法，即割集调节和树分解
- 启发式搜索中，评价函数的作用是：从当前节点出发来选择后续节点
- 启发式搜索中，启发函数的作用是：计算从当前节点到目标节点之间的最小代价值
- 贪婪最佳优先搜索中，评价函数取值=启发函数取值
- A*算法是一种有信息搜索算法，在最短路径搜索中引入的辅助信息是：任意一个城市到目标城市之间的直线距离
- 为了保证 A*算法是最优的，需要启发函数具有可容和一致的特点，下面对启发函数具有可容性这一特点的解释正确的是：

启发函数不会过高估计从当前节点到目标节点之间的实际开销代价
- 下面对 Alpha-Beta 剪枝搜索算法描述中正确的是：

节点先后次序对剪枝效率有影响

剪枝对算法输出结果没有影响

大多情况下剪枝会提高算法效率
- 下面对 minimax 搜索算法描述中正确的是：

需要遍历游戏树中所有节点

MAX 节点希望自己收益最大化

给定一个游戏搜索树，minimax 算法通过每个节点的 minimax 值来决定最优策略

MIN 节点希望对方收益最小化
- Alpha 和 Beta 两个值在 Alpha-Beta 剪枝搜索中被用来判断某个节点的后续节点是否可被剪枝，下面对 Alpha 和 Beta 的初始化取值描述正确的是：

Alpha 和 Beta 初始值分别为负无穷大和正无穷大

- 下面对上限置信区间 (UCB)算法在多臂赌博机中的描述，哪句描述是不正确的 (C) _

UCB 算法在探索-利用之间寻找平衡

UCB 算法既考虑拉动在过去时间内获得最大平均奖赏的赌博机，又希望去选择那些拉动臂膀次数最少的赌博机

UCB 算法每次随机选择一个赌博机来拉动其臂膀 (X)

UCB 算法是优化序列决策问题的方法

- 下面对 minimax 搜索、alpha-beta 剪枝搜索和蒙特卡洛树搜索的描述中正确的是： _

minimax 是穷举式搜索

对于一个规模较小的游戏树，alpha-beta 剪枝搜索和 minimax 搜索的结果相同

三种搜索算法中，只有蒙特卡洛树搜索是采样搜索

alpha-beta 剪枝搜索和蒙特卡洛树搜索都是非穷举式搜索

- 如果问题存在最优解，宽度优先搜索必然可以得到该最优解
- 如果问题存在最优解，启发式搜索可以认为是“智能程度相对较高”的算法
- 在启发式搜索策略中，closed 表用于存放已经扩展过的节点
- 在估价函数中，对于 $g(x)$ 和 $h(x)$ 下面描述正确的是： _

$h(x)$ 是从节点 x 到目标节点的最优路径的估计代价

$g(x)$ 是从初始节点到节点 x 的实际代价

- 遗传算法主要借用生物进化中“适者生存”的规律
- 遗传算法的适应度函数是用来区分群体中的个体好坏的标准
- 遗传算法中起核心作用的是交叉算子
- 遗传算法采用群体搜索策略，同时对搜索空间中的多个解进行评估，因此遗传算法具有较好的全局搜索性能

- 遗传算法不能保证每次都得到全局最优解
- 生物进化过程中选择通过遗传和变异起作用，同时又使变异和遗传向着适应环境方向发展
- 选择是通过遗传和变异起作用的，变异为选择提供资料，遗传巩固与积累选择的资料，而选择能控制变异与遗传的方向，使变异和遗传向着适应环境的方向发展
- 在遗传算法中，将所有妨碍适应度值高的个体产生，从而影响遗传算法正常工作的问题统称为欺骗问题
- 在遗传算法应用中，适应度函数的设计要结合问题本身的要求而定，但适应度函数和问题的目标函数没有关系
- 在遗传算法中，适应度大的个体被选择的概率大，但不是说一定能够被选上
- 关于蚁群算法，下面叙述正确的是：

蚁群算法是通过人工模拟蚂蚁搜索食物的过程，即通过个体之间的信息交流与相互协作最终找到从蚁穴到食物源的最短路径的

蚁群算法是一种应用于组合优化问题的启发式搜索算法

蚂蚁系统是一种增强型学习系统

- 关于蚁群算法的参数，下面叙述正确的是

信息素启发因子越小，蚁群搜索的随机性越小

信息素启发因子越大，蚂蚁选择以前走过的路径的可能性越大，蚁群的搜索过程越不易陷入局部最优

- 蚂蚁在运动过程中，根据各条路径上的信息素决定转移方向

- 什么是搜索？

利用已有知识和经验，根据问题的实际情况，不断寻找可利用知识，从而构造一条代价最小的推理路线，使问题得以解决的过程称为搜索

- 从初始节点 S_0 开始逐层向下扩展，在全部搜索完第 k 层节点之后，才进入第 $k+1$ 层节点进行搜索。这种搜索策略属于宽度优先搜索

- 启发性信息包括（ ）有关的信息

与具体问题求解过程

指导搜索过程

搜索最有希望方向

- 启发性信息的作用强度会影响搜索结果，当强度为强时，可降低搜索工作量，但可能导致 找不到最优解

- 双人完备信息博弈是指两位选手对垒，轮流走步，双方都知道（ ）

对方已经走过的棋步

估计出对方未来的走步

棋局的初始状态

- 博弈树的搜索方法采用的是

极大极小分析法

- 设二方博弈已生成一棵博弈搜索树，一方寻找最优行动方案，需进行（ ）等

推算父节点的得分

设计一个估价函数

估算当前节点的得分

- 对于一棵博弈搜索树，自己方寻找最优行动方案时，取估价值的原则是：估价函数取正值

- 盲目搜索的搜索策略包括（ ）优先搜索

深度、广度

- 以下不是搜索问题一般具有的特征是：

问题一定是有解的

- 在搜索过程中，当扩展节点时，以下说法错误的是 D

一个节点经一个算符操作通常只生成一个子节点

扩展节点时，会生成该节点的所有后继节点

判断子节点是否是当前扩展节点的父节点、祖父节点等，若是，则删除。

通常适用于一个节点的算符只有一个 (X)

- 普通搜索问题就是指：

求出一条从初始状态到目标状态之间的行动序列问题

- 盲目搜索错误的是：

盲目搜索不需要重排 OPEN 表

盲目搜索按预定的控制策略进行搜索，在搜索过程中获得的中间信息不用来改进控制策略 (by南大ai院小妖兽，盗版必究)

盲目搜索通常都会找到最优解 (X)

盲目搜索不便于解决复杂问题

- 问题求解的目的包括：

希望机器找到问题的最优解

希望机器找到问题的一个解

希望机器能够举一反三，具有解决同类问题的能力

- 搜索问题通常包括以下哪些重要组成因素：

后继函数、初始状态、路径代价函数

- 以下那种情况优先选择广度优先搜索

解出现在相对较浅的位置

节点分支因子不是太大

- 问题求解技术是从人工智能初期的智力难题、棋类游戏、简单数学定理证明等研究中发展起来的

- 在状态空间图中，用节点表示状态，用无向边表示算符

- 如果我们要求所得的操作序列能够使得总代价最低，这样的搜索问题称为最优搜索问题

- 当前状态达到目标状态时，由初始状态到目标状态所使用的算符组成的序列就是问题的一个解

- 搜索分为盲目搜索和启发式搜索

- 盲目搜索策略包括：广度优先、有界深度优先、深度优先

- 盲目搜索不包括全局择优搜索

- 搜索过程中，当扩展节点时，以下说法错误的是：

一个节点经过一个算符操作通常只生成一个子节点

通常适用于一个节点的算符只有一个 (X)

判断子节点是否是当前扩展节点的父节点、祖父节点等，若是，则删除

扩展节点时，会生成该节点的所有后继节点

- 搜索问题通常包括以下哪些重要组成因素：后继函数、初始状态、目标测试

- 关于贪心算法，下面说法正确的是：

贪心算法可以快速地找到一个可行解，但不一定找到最优解

贪心算法也是先将一个问题分成几个步骤操作

贪心算法在每一步选择当前看起来是最佳的选择

- 实现搜索所需要的数据结构中，下面说法正确的是：

不同的搜索策略，节点在 OPEN 表中的排列顺序是不同的

不同的搜索策略，节点在 CLOSED 表中的排列顺序是相同的

- 对问题的求解，是人工智能要解决的核心问题之一

- 以下不是搜索问题一般具有的特征是：

问题一定是有解的

- 关于盲目搜索错误的是：D

盲目搜索不需要重排 OPEN 表

盲目搜索按预定的控制策略进行搜索，搜索过程中获得的中间信息不用来改进控制策略

盲目搜索不便于解决复杂问题

盲目搜索通常会找到最优解 (X)

- 下列关于迭代加深的深度优先搜索的说法错误的是：

迭代加深的深度优先结合了广度优先和深度优先搜索两者的优点

深度优先搜索最坏情况下的时间复杂度是指数级的，但迭代加深的深度优先搜索最坏情况下的时间复杂度不是指数级的 (X)

迭代加深的深度优先搜索是完备的

迭代加深的深度优先搜索是最优的

- 搜索问题通常要包括以下哪些重要的组成要素：_

目标测试、后继函数、初始状态、路径代价函数

- 在生成与测试范式中，对于一个好的生成器，他应该能够做到：

生成器应该是完备的、知情的、非冗余的

- 对问题的求解是人工智能要解决的核心问题之一
- 人工智能虽然有多个应用领域，而且每个应用领域又各有自己的规律和特点，但从他们求解具体问题的过程来看，都可以抽象为一个问题求解的过程
- 估价函数一般形式为： $f(n)=g(n)+h(n)$ ，其中 $h(n)$ 称作：启发函数
- 关于爬山法的说法中错误的是：

爬山法是一种贪心算法

爬山法是一种局部择优的方法

爬山法没有能力从错误中或错误路径中恢复

爬山法需要保存未选择路径的记录 (X)

- A*搜索可以认为是“智能程度比较高”的算法
- 下面关于启发式搜索正确的是 (by南大ai院小妖兽盗版必究)

在启发式搜索中，对节点的评价是十分重要的，评价函数是搜索成败的关键

启发式搜索可以省略大量无谓的搜索路径

在搜索过程中对待扩展的每一个节点进行评估，得到最好的位置，再从这个位置进行搜索直到目标

启发式搜索，也称为有信息搜索或知情搜索，借助问题的特定知识来帮助选择搜索方向

- 下面关于基本爬山法的说法中错误的是：

只估计剩余距离，而忽略了实际走过的距离

在给定的状态下做出决策，并且从多个可能的选项中做出最好的决定 (X)

每次拿相邻点与当前点进行比对，取两者中的较优者，作为爬坡的下一步

没有真正足够的信息确定自己在正确的路径上

- 关于最佳优先搜索，下面说法正确的是：

最佳优先搜索实现也需要 open 表和 closed 表

Open 表中节点按照节点接近目标状态的启发式估计值进行顺序排列

最佳优先搜索是智能搜索算法

- 一定能够找到最优解的启发函数是包括

分支定界法

A*搜索算法

- 关于 A*算法，以下说法正确的是

A*算法是一个应用非常广泛的算法

A*算法按照路径长度对 open 表中节点进行排序

A*算法结合了具有剩余距离估计值和动态规划的分支定界法

- 启发式搜索下列错误的是：

启发式搜索是一个提高复杂问题解决效率的搜索策略

启发式函数可以极大地减少所搜索的节点数目

启发式搜索根据启发式估计值指引搜索方向，如果存在解，一定能找到解 (X)

启发式搜索沿着一条最可能的路径到达解，忽略最没有希望的路径

- 关于与/或树中的说法中，正确的有：

与/或树通常用于复杂问题的求解

与/或树是一种用于问题简化的技术

与/或树是用于表示问题及其求解的过程

- 爬山法可能出现的问题有：

高原问题、山脊问题、山麓问题

- 关于使用低估值的分支定界法，说法正确的是：

使用低估值的分支定界法，会利用启发信息对节点到达目标节点的值进行估计

使用低估值的分支定界法，按照估计的总长度的代价来生成路径

与普通分支定界法不同的是，扩展节点时，需要估算法当前节点的每个子节点到目标节点的距离，加上当前节点的路径长度，作为新的路径长度

无信息搜索的总结

(b: 最大分支个数, d: 解所在的层数, C^* : 最优解代价, ε : 最小单步代价值)

(m: 树的高度 l: 深度限制)

	BFS	一致代价	DFS	深度有限	迭代深入	minimax
完备性	yes	yes	no	no	yes	yes
时间复杂度	$O(b^{d+1})$	$O(b^{\lceil C^*/\varepsilon \rceil})$	$O(b^m)$	$O(b^l)$	$O(b^d)$	$O(b^m)$
空间复杂度	$O(b^{d+1})$	$O(b^{\lceil C^*/\varepsilon \rceil})$	$O(bm)$	$O(bl)$	$O(bd)$	$O(bm)$
最优性	Yes	yes	No	No	Yes	Yes(仅对手最优时)

迭代深入搜索的性质中，最优性可以保证 \Leftrightarrow 单步代价相等

松弛问题

- 对原问题的动作的约束放宽，成为问题的松弛化
- 松弛问题的最优解的代价就可以用来定义原问题的一个可采纳的启发式函数

模拟退火搜索的性质

- 如果 T 降低的足够慢，则模拟退火能以趋于 1 的概率找到最优解
- 广泛应用于 VLSL 布局问题，航空调度问题

一致性代价搜索

- 完备性：yes \Leftrightarrow 单步代价不是无穷小
- 时间：代价小于最优解的结点个数
- 空间：代价小于最优解的结点个数
- 最优性：结点是根据代价排序拓展的

启发式搜索保证最优性的条件

- 可采纳式：不会过估计到达目标的代价
- 一致性： $\forall n, n'$, n' 是 n 的后继，从 n 到目标的代价不大于从 n' 到目标的代价加上从 n 到 n' 的代价
- $h(n)$ 是可采纳的 \Rightarrow 树搜索最优
- $h(n)$ 是一致的 \Rightarrow 图搜索最优

爬山法的局限性

- 局部最优
- 山脊：造成一系列的局部最优
- 高原：平行坐标轴，也叫山肩

随机爬山法：在上山过程中随机选择下一步，选中结点的概率和导数（变化率，陡峭程度）

有关

首次爬山法：随机生成一个后继结点

随机重启爬山法：随机生成初始状态，多次尝试，成功概率依概率收敛到 1

GSP 的回溯搜索

- GSP 问题与变量赋值的顺序无关
- 考虑每次只给一个变量赋值

如何提高 GSP 和效率

- 下一步给哪个变量赋值
- 以什么顺序给变量赋值
- 怎么提前找到所有的失败？

GSP 变量的取值顺序

- 先找具有最少合法赋值的变量
- 选择可以约束其他未赋值变量最多的变量
- 选择对其他未赋值变量影响最小的合法赋值—赋值原则

} 选择变量的原则

前向检测

- 每个变量建立一张表，记录合法赋值
- 检查与 x 相关的未赋值变量 y ，从 y 中删除与 x 不相容的值
- 如果发现有某个变量的合法赋值为空，则回溯
- 缺点：不能提早检测出所有的失败

结点相容：CSP 网络中的结点值域中所有取值满足它的一元约束

弧相容 \Leftrightarrow 对 x 的每一个变量， y 都存在一个合法赋值，则 $x \rightarrow y$ 弧相容

- 删除不满足弧相容的合法赋值
- 如果 x 删除了一个合法赋值，则对他的邻居需重新检查，邻居 $\rightarrow x$ 是否弧相容

路径相容：通过观察变量得到隐式约束，从而加强弧相容

k 相容：如果任何 k-1 个变量相容，那么第 k 个变量的赋值也能与前 k-1 个变量相容，那么
这个 csp 就是 k 相容的

博弈搜索

- 博弈通常指：确定的、完全可观察的环境中两个 agent 必须轮流行动的零和游戏
- 博弈中如何做出决策？极小极大值算法：选择具有最高极小极大值的动作来执行

$$\text{Mini max}(s) = \begin{cases} \text{Utility}(s) & s \text{ 为终止状态} \\ \max_{a \in \text{Action}(s)} \text{Mini max}(\text{Result}(s, a)) & s \text{ 为 max 结点} \\ \min_{a \in \text{Action}(s)} \text{Mini max}(\text{Result}(s, a)) & s \text{ 为 min 结点} \end{cases}$$

- 极小极大值定义：

解释：如果当前为终止状态，返回效能函数

如果当前为 max 结点，返回它后续动作的最大值

如果当前为 min 结点，返回它后续动作的最小值

- 效能函数 Utility：评估终止状态，取值 -1, 0, 1
- 极小极大值的计算，不需要遍历博弈树中的每个节点
- 极小极大值搜索的游戏状态数目，随着搜索层指数级增长
- α - β 剪枝可以尽可能消除搜索树

α ：max 路径上至今为止发现的最大值（最好结果）

如果 v 比 α 差，max 会避免它，发生剪枝

β ：min 路径上的最小值（最好结果）

问题形式化的四个部分

- 初始状态
- 动作或后续函数
- 目标测试
- 路径代价 (注：一个解是从初始状态到目标状态的动作序列)

逻辑推理

- 知识是根据信息来定义的
- 过程性知识用于描述问题如何求解
- 陈述性方法可以使用一阶逻辑
- 下列关于语义网络正确的陈述：

当所具有的知识可被很好地理解为一组彼此相关的概念时，可以使用语义网络

可表示为一种直接或间接的图，由表示概念的节点和表示概念之间语义关系的弧组成。

可以表示概念之间的语义关系

- 本体提供一个领域的公共词汇，并且定义一些术语的含义和它们之间的关系
- 本体是一种对特定论域中实体的类型、特性和相互关系的形式化命名和定义
- 贝叶斯网络又称信念网络或因果网络
- 知识库系统由知识库和搜索引擎组成，其中，知识库表示关于世界的事实，推理引擎则可以

对这些事实进行推理

- 知识表示关注于设计计算机表示来采集关于世界的知识，可用于解决复杂的问题
- 多效用器和多躯体这两者都是单智能体问题
- 经典规划的特性：

一个智能体仅有一个已知的初始状态

每次保持确定性的动作

静态环境中部分可观测（错）

动态环境中完全可观测（错）（不知道对的是啥）

- 可用于经典规划的方法有：

约束满足、状态空间搜索、布尔可满足性、规划精进、一阶逻辑推理

- 现实世界中用于规划和调度的规划者更为复杂，可选择表示语言、与环境交互的方式扩展
- 多智能体的特征是：分散化、自主性
- 要仿真鸟群的群体行为，其规则应该包括：对齐、分离
- 有两种搜索计划的方式：前向状态空间搜索、后向状态空间搜索
- 马尔可夫决策过程（MDP）是一种离散时间随机控制过程，意味着动作结果仅仅依赖于

当前状态

- 动态规划中有两种优化策略：值迭代、策略迭代
- 如果命题 p 为真、命题 q 为假，则下述哪个复合命题为真命题

如果 q 则 p

- p 和 q 均是原子命题，“如果 p 那么 q ”是由 p 和 q 组合得到的复合命题。下面对“如果 p 那么 q ”这一复合命题描述不正确的是 (A)

无法用真值表来判断“如果 p 那么 q ”的真值 (X)

“如果 p 那么 q ”定义的是一种蕴涵关系（即充分条件）

“如果 p 那么 q ”意味着命题 q 包含着命题 p ，即 p 是 q 的子集

当 p 不成立时，“如果 p 那么 q ”恒为真

- 下面哪个复合命题与“如果秋天天气变凉，那么大雁南飞越冬”是逻辑等价的

如果大雁不南飞越冬，那么秋天天气没有变凉

- 下面哪一句话对命题范式的描述是不正确的 (C)

有限个简单合取式构成的析取式称为析取范式

有限个简单析取式构成的合取式称为合取范式

一个析取范式是不成立的，当且仅当它包含一个不成立的简单合取式 (X)

一个合取范式是成立的，当且仅当它的每个简单析取式都是成立的

- 下面哪个逻辑等价关系是不成立的 (A)

$$\exists x P(x) \equiv \neg \forall x P(x) \quad (X)$$

$$\forall x P(x) \equiv \neg \exists x \neg P(x)$$

$$\forall x \neg P(x) \equiv \neg \exists x P(x)$$

$$\neg \forall x P(x) \equiv \exists x \neg P(x)$$

- 下面哪个谓词逻辑的推理规则是不成立的 (D)

存在量词引入(Existential Generalization, EG): $A(c) \rightarrow (\exists x)A(x)$

全称量词消去(Universal Instantiation, UI): $(\forall x)A(x) \rightarrow A(y)$

存在量词消去(Existential Instantiation, EI): $(\exists x)A(x) \rightarrow A(c)$

全称量词引入(Universal Generalization, UG): $(\exists y)A(y) \rightarrow (\forall x)A(x)$

- 下面对知识图谱描述正确的是:

知识图谱中一条边可以用一个三元组来表示

知识图谱中的节点可以是实体或概念

知识图谱中一条边连接了两个节点, 可以用来表示这两个节点存在某一关系

知识图谱中两个节点之间可以存在多条边

- 如果知识图谱中有 David 和 Mike 两个节点, 他们之间具有 classmate 和 brother 关系。

在知识图谱中还存在其他丰富节点和丰富关系 (如 couple, parent 等) 前提下, 下面描述正确的是:

可以从知识图谱中形成 classmate<David, Mike>的表达, 这里 classmate 是谓词。

可从知识图谱中找到 classmate 和 brother 这个谓词的正例和反例。

可以从知识图谱中形成 brother<David, Mike>的表达, 这里 brother 是谓词。

- 在一阶归纳学习中，只要给定目标谓词，FOIL 算法从若干样例出发，不断测试所得到推理规则是否还包含反例，一旦不包含负例，则学习结束，展示了“归纳学习”能力。下面所列出的哪个样例，不属于 FOIL 在学习推理规则中所利用的样例的是 (D)

背景知识样例

所得到推理规则涵盖的正例

所得到推理规则涵盖的反例

信息增益超过一定阈值推理规则涵盖的例子 (X)

- 对一阶归纳推理 (FOIL) 中信息增益值(information gain)阐释正确的是：信

息增益值大小与背景知识样例数目无关。

在计算信息增益值过程中，需要利用所得到的新推理规则和旧推理规则分别涵盖的正例和反例数目。

在算法结束前，每次向推理规则中加入一个前提约束谓词，该前提约束谓词得到的新推理规则具有最大的信息增益值。

信息增益值用来判断向推理规则中所加入前提约束谓词的质量。

- 基于知识图谱的路径排序推理方法可属于如下哪一种方法： 监督学习
- 下面哪个步骤不属于基于知识图谱的路径排序推理方法中的一个步骤：

定义和选择若干标注训练数据

定义和选择训练数据的特征

定义和选择某一特定的分类器

定义和选择 do 算子操作 (X)

- 在基于知识图谱的路径排序推理方法中，最后训练所得分类器的功能是

给定两个实体（知识图谱中的两个节点），判断其是否具有分类器所表达的关系（即节点之间是否具有分类器所能够辨认的关系）

- 下面对辛普森悖论描述不正确的是：A

辛普森悖论的原因在于数据之间相互不关联。（X）

如果忽略一些潜在因素，可能会导致全部数据上观察到的结果却在部分数据上不成立。

为了克服辛普森悖论，需要从观测结果中寻找引发结果的原因，由果溯因。

在某些情况下，忽略潜在的“第三个变量”，可能会改变已有的结论，而我们常常却一无所知。

- 下面哪个模型不属于用来进行因果推理的模型：D

因果图

结构因果模型

贝叶斯网络

有向有环图（X）

- 下列正确的是：

在有向无环图中，父辈节点“促成”了孩子节点的取值

一个有向无环图唯一地决定一个联合分布

一个联合分布不能唯一地决定有向无环图

- 因果推理中，D-分离的作用是

判断集合 A 中的变量是否与集合 B 中的变量相互独立

- 下面对干预和 do 算子描述不正确的是：C

do(x)=a 表示将 DAG 中指向节点 x 的有向边全部切断，并且将 x 的值固定为常数 a

干预指的是固定(fix)系统中某个变量，然后改变系统，观察其他变量的变化

do(x)=a 表示将 DAG 中指向节点 x 的所有节点取值均固定为 a (X)

do 算子的意思可理解为 “干预”

- 对反事实推理描述正确的是

反事实推理是用于因果推理的一种方法。

条件变量对于结果变量的因果性就是A 成立时 B 的状态与 A 取负向值时 “反事实” 状态 (B') 之间的差异。如果这种差异存在且在统计上是显著的，说明条件变量与结果变量存在因果关系。

事实是指在某个特定变量(A)的影响下可观测到的某种状态或结果(B)。“反事实” 是指在
该特定变量(A)取负向值时可观测到的状态或结果(B')。

- 下面对通过因果图来进行因果推理所存在不足描述不正确的是：A

因果图无法刻画数据之间的联合分布。(X)

难以得到一个完整的DAG 用于阐述变量之间的因果关系或者数据生成机制，使得 DAG 的应用受到的巨大的阻碍。

DAG 作为一种简化的模型，在复杂系统中可能不完全适用，需要将其拓展到动态系统 (如时间序列)。

在因果推理中引入了 do 算子，即从系统之外人为控制某些变量。但是，这依赖于一个假定：干预某些变量并不会引起 DAG 中其他结构的变化。

- 一个命题不能同时既为真又为假，但可以在一种条件下为真，在另一种条件为假

- 李明的父亲是教师，用谓词逻辑可以表示为 $\text{Teacher}(\text{father}(\text{Liming}))$ 这里 $\text{father}(\text{Liming})$ 是：函数

- 命题逻辑可以把所描述的事物的结构及其逻辑特征反映出来，也能把不同事物间的共同特征表述出来 (X)

- 表示“每个人都有喜欢的人”的是：

$$(\forall x)(\exists y)\text{Like}(x, y)$$

- 对于谓词公式 $\exists x(P(x, y) \rightarrow Q(x, y)) \vee R(x, y)$ ，以下说法错误的是：C

上述公式中的所有 y 是自由变元

$P(x, y)$ 中的 x 是约束变元

$R(x, y)$ 中的 x 是约束变元 (X)

$Q(x, y)$ 中的 x 是约束变元

- 一阶谓词逻辑表示的优点是

自然性、精确性、严密性、易实现

- 下列属于谓词公式的是：

$P(x)$

非 $P(x)$

$P(x) \leftrightarrow Q(x)$

- 一阶谓词逻辑表示法可以表示不确定的知识 (X)
- 产生式是蕴含式 (X)
- 不适合产生式表示法的知识是：具有结构关系的知识

- 适合产生式表示法表示的知识是：

由许多相对独立的知识元组成的领域知识

可以表示为一系列相对独立的求解问题的操作

具有经验性及不确定性的知识

- 框架表示法可以表示具有因果关系的知识
- 产生式系统求解问题的过程是一个反复进行“匹配--冲突消解--执行”的过程
- 模块性不是框架表示法特点
- 下列是框架表示法特点的是
 - 结构性、继承性、自然性
- 框架的槽值或侧面值可以是另一个框架的名字
- 一个产生式系统由规则库、推理机、综合数据库三部分组成
- 产生式有固定的格式，每一条产生式规则都由前提与结论（操作）两部分组成
- 任何文字的合取式称为子句
- 空子句是无法满足的
- 谓词公式不可满足的充要条件是其子句集不可满足
- 对于一阶谓词逻辑，若子句集是不可满足的，必存在一个从该子句集到空子句的归结演绎
- 对于一阶谓词逻辑，如果没有归结出空子句，则说明原谓词公式是不可满足的
- 如果证据 E 的出现使得结论 H 一定程度为真，则可信度因子 $0 < CH(H, E) < 1$
- 在可信度方法中，若证据 A 的可信度 $CF(F) = 0$, 这意味：对证据一无所知
- 在证据理论中，信任函数与似然函数对 $(Bel(A), Pl(A))$ 的值为 $(0, 0)$ 时，表示：A 为假

- 不确定推理中，除了需要解决推理方法、推理方向、控制策略外，还需解决：

组合证据不确定性的算法

结论不确定性的合成

不确定的表示与度量

不确定性的传递算法

- 基本概率分配函数之值是概率 (X)
- 若模糊推理结果为 $0.3/-4+0.8/-3+1.0/-2+1.0/-1+0.8/0+0.3/1+0.1/2$ ，根据最大隶属度平均法，模糊决策的结果为：-1.5

- 模糊性是由事物的概念界限模糊和人的主观推理与判断产生的
- 模糊集合与其隶属函数是等价的 (X)
- 隶属函数是对模糊概念的定量描述
- 隶属函数的确定带有主观性
- 二元模糊关系是指两个模糊集合的元素间所具有关系的程度
- 模糊推理是利用模糊性知识进行的一种不精确推理
- 专家系统与计算机程序的区别是：

专家系统研究的是符号表示的知识而不是数值数据为研究对象

专家系统的控制结构与知识是分离的

专家系统采用启发式搜索方法而不是普通的算法

- 专家系统的基本结构包括：人机界面、解释模块、知识库、推理机、知识库管理系统、动态数据库

- 专家系统的瓶颈是知识的获取
- 专家系统是一个大型的软件，它的功能应该不低于人类专家解决问题的能力

- 谓词公式可以表示事实性的知识例如事物的（ ）等

概念、属性、状态

- 谓词逻辑中使用的逻辑链接符包括

合取、析取、蕴含

- 集合之间的关系包括：

子集、相等、真子集

- 集合之间的运算包括：

相对补集、并集、交集（不包括对称）

- 将命题“存在最小的自然数”符号化，需要使用

全称和存在量词

- 模糊集合论适合于表示事物

状态的模糊性

- 隶属函数属于模糊集合理论范围的概念

- 经典概念有三个组成部分：

概念名、概念的内涵表示、概念的外延表示

- 逻辑命题中的复合命题是通过逻辑连接词将原子命题相互联结而成的

- 蕴含式的匹配总要求是精确的。

- 产生式匹配可以是精确的，也可以是不精确的，只要按照某种算法求出的相似度落在预先 指定的范围内，就可以说是可匹配的

- 可信度带有较大的主观性和经验性，其准确性难以把握

- 一个好的规则库具有以下哪些特征：

知识完整、知识一致、组织合理、表达灵活

- 以下属于控制系统工作原理的是：

按一定策略从规则库中选择规则与综合数据库中的已知事实进行匹配

随时掌握结束产生式系统运行的时机，以便在适当的时候停止系统的运行

匹配成功的规则可能不止一条，这称为发生了冲突。此时，推理机构必须调用相应的解决冲突的策略进行消解，以便从中选出一条执行。

在执行某一条规则时，如果该规则的右部是一个或者多个结论，那么把这些结论加入到综合数据库中；如果该规则的右部是一个或者多个操作，则执行这些操作

- 常见的不确定性合成方法有：

递推算法、求极大值法、加权求和法

- 综合数据库是存放问题求解过程中各种当前信息的数据结构
- 综合数据库中的已知事实通常使用字符串、向量、集合、矩阵、表 等数据结构表示
- 控制系统又称为推理机构，由一组程序组成，负责整个产生式系统的运行，实现对问题的求解
- 根据经验对一个事物或现象为真的相信程度称为可信度

学习

- 归纳学习用于从训练样本中发现一般的事实
- 有一组供用户标识为感兴趣或不感兴趣的网页，可被机器学习用来解决网页过滤问题，选择下列特征中哪个是通过训练这些网页学来的？经验
- CNN 是一种可用于从图像中识别视觉模式的深度神经网络
- 哪些是机器学习方法：决策树、自适应提升、随机森林、感知机
- 机器学习的三个派别是：联结主义、符号主义、行为主义
- 机器学习可以解决的任务有：分类、聚类、回归、降维
- 典型的机器学习范式是：无监督学习、有监督学习
- 机器学习是人工智能的一个分支
- 分类任务：为每个项目分配一个类别
- 回归任务：预测每个项目实际的值
- 排名任务：依据某个准则对项目进行排序
- 聚类任务：对数据对象进行分组
- 密度估计任务：发现某个空间中输入的分布
- 降维任务：将高维空间映射到低维空间来简化输入
- 回归的输出是一个实际的连续值，分类的输出是一个离散类别
- 聚类是为输入对象标识相似的组，没有训练数据；而分类是给输入项分派预定义的类，有训练数据
- 有监督学习：接收一组标注数据并对所有未知数据进行预测
- 半监督学习：属于有监督学习算法一类，此外还利用未标记数据进行训练
- 无监督学习：仅仅接收未标注数据，并对所有未知数据进行预测

- 弱监督学习：旨在采用有限数量的训练样本来学习一些信息
- 一次性学习：旨在从一个、或仅有的几个训练样本中学习一些信息
- 零次性学习：能够求解一个任务，即使没有得到该任务的任何训练样本
- 迁移学习把储备的强知识用于不同但相关的问题
- 下列哪些算法已被发明为专门用于强化学习算法：

异步优势动作者•评论者

神经情景控制

确定性策略梯度

- 已被用作机器学习的代表性模型有

概率模型、逻辑模型、几何模型、网络化模型

- 已经被用作机器学习的概率模型有

概率规划、隐马尔可夫模型、朴素贝叶斯和贝叶斯网络、高斯过程

- 真正被流形学习所使用的算法有：

LLE（局部线性嵌入）、Isomap（等距映射）、LE（拉普拉斯特征映射）

- 真正被用作机器学习逻辑模型的方法：

关联规则、决策树、一阶逻辑

- 真正用于机器学习的人工神经网络有：

卷积神经网络、长短期记忆、深度自动编码器

- 概率模型是采用概率论来表示所有不确定性的格式

- 几何模型直接在实例空间构建，该空间可被认为是几何的概念，例如：欧几里德空间或黎曼空间

- 人工神经网络是一种人脑的人工表示，试图仿真其学习过程

- 下面对经验风险和期望风险的描述中，哪一个属于过学习：

经验风险小，期望风险大

- 下面对结构风险最小化的描述中，哪一个描述是不正确的：C

结构风险最小化与经验风险最小化的目标是不同的

结构风险最小化在最小化经验风险与降低模型复杂度之间寻找一种平衡

为了更好地保证结构风险最小化，可适当减少标注数据 (X)

在结构风险最小化中，优化求解目标为使得经验风险与模型复杂度之和最小

- 监督学习方法又可以分为生成方法与判别方法，下面哪个不属于判别方法：B

Ada boosting

贝叶斯方法 (X)

回归模型

神经网络

- 下面哪句话语较为恰当刻画了监督学习方法中生成方法的特点：C

毕其功于一役

三个臭皮匠、抵一个诸葛亮

授之于鱼、不如授之于“渔”

屡败屡战、屡战屡败、最后成功

- 线性回归模型中，所优化的目标函数是：最小化残差平方和的均值

- 下面对 Ada Boosting 描述不正确的是：C

该算法将若干弱分类器线性加权组合起来，形成一个强分类器

在所构成的强分类器中，每个弱分类器的权重是不一样的

在所构成的强分类器中，每个弱分类器的权重累加起来等于 1 (X)

在每一次训练弱分类器中，每个样本的权重累加起来等于 1

- 下面对回归和分类描述不正确的是：B

两者均是学习输入变量和输出变量之间潜在关系模型

回归是一种无监督学习、分类学习是有监督学习 (X)

在分类模型中，学习得到一个函数将输入变量映射到离散输出空间

在回归分析中，学习得到一个函数将输入变量映射到连续输出空间

- 由于 K 均值聚类是一个迭代过程，我们需要设置其迭代终止条件。下面哪句话正确描述了 K 均值聚类的迭代终止条件：

已经达到了迭代次数上限,或者前后两次迭代中聚类质心基本保持不变

- 我们可以从最小化每个类簇的方差这一视角来解释 K 均值聚类的结果，下面对这一视角描述不正确的是：B

每个样本数据分别归属于与其距离最近的聚类质心所在聚类集合

每个簇类的质心累加起来最小 (X)

最终聚类结果中每个聚类集合中所包含数据呈现出来差异性最小

每个簇类的方差累加起来最小

- K 均值聚类算法的不足有：

算法迭代执行

需要初始化聚类质心

需要事先确定聚类数目

- 下列对于数据样本方差解释正确的是：

方差（样本方差）是每个样本数据与全体样本数据平均值之差的平方和的平均数

方差刻画了随机变量或一组数据离散程度

方差描述了样本数据在平均值附近的波动程度

- 协方差可以用来计算两个变量之间的相关性，或者说计算两维样本数据中两个维度之间的相关性。基于计算所得的协方差值，可以来判断样本数据中两维变量之间是否存在关联关系。 下列不正确的是：B

当协方差值大于 0 时，则两个变量线性正相关

当协方差值等于 0 时，则两个变量线性正相关 (X)

当协方差值小于 0 时，则两个变量线性负相关

当协方差值等于 0 时，则两个变量线性不相关

- 皮尔逊相关系数可将两组变量之间的关联度规整到一定的取值范围内。下面对皮尔逊相关系数描述不正确的是 (B)

皮尔逊相关系数的值域是 $[-1,1]$

- 皮尔逊相关系数刻画了两组变量之间线性相关程度，如果其取值越大，则两者在线性相关的意义下相关程度越大；如果其值越小，表示两者在线性相关的意义下相关程度越小 (X)

皮尔逊相关系数是对称的

皮尔逊相关系数等于 1 的充要条件是该两组变量具有线性相关关系

- 下面对相关性和独立性(independence)描述不正确的是：A

“不相关”是一个比“独立”要强的概念，即不相关一定相互独立

如果两维变量彼此独立，则皮尔逊相关系数等于 0

独立指两个变量彼此之间不相互影响

如果两维变量线性不相关，则皮尔逊相关系数等于 0

- 假设原始数据个数为 n ，原始数据维数为 d ，降维后的维数为 l ，下面对主成分分析算法描述不正确的是 (D)

在主成分分析中，我们将带约束的最优化问题，通过拉格朗日乘子法将其转化为无约束最优化问题

每个原始数据的维数大小从 d 变成了 l

主成分分析要学习一个映射矩阵，其大小是 $d \times l$

主成分分析学习得到了 l 个 d 维大小的向量，这 l 个 d 维向量之间彼此相关 (X)

- 在主成分分析中，我们将带约束的最优化问题，通过拉格朗日乘子法将其转化为无约束最优化问题。下面对主成分分析中优化的目标函数和约束条件描述正确的是：

保证映射投影（即降维后）所得结果方差最大以及投影方向正交（以去除冗余度）

- 下面对特征人脸算法描述不正确的是（ A ）

特征人脸之间的相关度要尽可能大

特征人脸方法是用一种称为 "特征人脸(eigenface)" 的特征向量按照线性组合形式来表达每一张原始人脸图像

每一个特征人脸的维数与原始人脸图像的维数一样大

特征人脸方法是一种应用主成分分析来实现人脸图像降维的方法

- 下面对逻辑斯蒂回归(logistic regression)描述不正确的是（ C ）

逻辑斯蒂回归中所使用 Sigmoid 函数的输出形式是概率输出

逻辑斯蒂回归是一种非线性回归模型

在逻辑斯蒂回归中，输入数据特征加权累加值在接近 $-\infty$ 或 $+\infty$ 附近时，模型输出的概率值变化很大

在逻辑斯蒂回归中，能够实现数据特征加权累加

- logistic 回归属于监督学习

- 潜在语义分析属于无监督学习

- 线性区别分析属于监督学习

- logistic 回归函数也被称作对数几率回归

- 在隐性语义分析中,给定 M 个单词和 N 个文档所构成的单词-文档矩阵(term-document)

矩阵,对其进行分解,将单词或文档映射到一个 R 维的隐性空间。描述不正确的是 (B)

单词和文档映射到 R 维隐性空间后,单词和文档具有相同的维度

这一映射过程中需要利用文档的类别信息

通过矩阵分解可重建原始单词-文档矩阵,所得到的重建矩阵结果比原始单词-文档矩阵更好捕获了单词-单词、单词-文档、文档-文档之间的隐性关系

隐性空间维度的大小由分解过程中所得对角矩阵中对角线上不为零的系数个数所决定

- 线性区别分析(linear discriminant analysis, LDA)在进行数据降维时,原始高维数据被映射到低维空间中后需要达到的优化的目标是:

既要保证同一类别样本尽可能靠近,也要保证不同类别样本尽可能彼此远离

- LDA(线性区别分析)与 PCA (主成分分析)均是降维的方法,下面描述不正确的是 (B)

PCA 对高维数据降维后的维数是与原始数据特征维度相关 (与数据类别标签无关)

PCA 和 LDA 均是基于监督学习的降维方法

假设原始数据一共有 K 个类别,那么 LDA 所得数据的降维维度小于或等于 K-1

LDA 降维后所得维度是与数据样本的类别个数 K 有关 (与数据本身维度无关)

- 下面对逻辑斯蒂回归 (logistic regression)和多项逻辑斯蒂回归模型(multi-nominal logistic model)描述不正确的是 (C)

多项逻辑斯蒂回归模型也被称为 softmax 函数

两者都是监督学习的方法

逻辑斯蒂回归是监督学习,多项逻辑斯蒂回归模型是非监督学习

两者都可被用来完成多类分类任务

- 逻辑斯蒂回归和线性区别分析均可完成分类任务，下面描述正确的是

逻辑斯蒂回归可直接在数据原始空间进行分类，线性区别分析需要在降维所得空间中进行分类

- 几种学习的学习方式：

强化学习：评估学习方式

监督学习：有标注信息学习方式

深度卷积神经网络学习：端到端学习方式

- 在强化学习中，通过哪两个步骤迭代，来学习得到最佳策略：策略优化与策略评估
- 在强化学习中，哪个机制的引入使得强化学习具备了在利用与探索中寻求平衡的能力：

贪心策略

- 深度强化学习是深度学习与强化学习的结合。在深度强化学习中，神经网络被用来进行哪个函数的学习

q 函数

- 与马尔科夫奖励过程相比，马尔科夫决策过程引入了哪一个新元素：动作
- 在本课程所讲述的范围内，“在状态 s ，选择一个动作，使得状态 s 得到最大的反馈期望”，

这句话描述了状态 s 的：策略学习与优化

- 在本课程内容范围内，“在状态 s ，按照某个策略行动后在未来所获得反馈值的期望”，这句话描述了状态 s 的：价值函数

· 在本课程内容范围内，“在状态 s ，按照某个策略采取动作 a 后在未来所获得反馈值的期望”，这句话描述了状态 s 的：动作-价值函数

- 下列哪句话描述了马尔科夫链中定义的马尔科夫性： $t+1$ 时刻状态取决于 t 时刻状态

· 下面对强化学习、有监督学习和无监督学习描述正确的是：都是人工智能的学习算法

· 下列属于机器学习研究内容的是

学习方法、学习机理、学习系统（学习对象不属于）

· 下列属于机器学习所涉及学科的是

逼近论、统计学、概率论（心理学不属于）

· 下列选项属于机器学习主要策略是

解释学习、示教学习、机械学习（演绎学习不属于）

by南大ai院小妖兽

导论

- 图灵测试旨在给予哪一种令人满意的操作定义？机器智能

- 人工智能概念正确的是：

人工智能旨在创造智能机器

人工智能是研究和构建在给定环境下表现良好的智能体程序

人工智能是通过机器或软件展现的智能

- 以下学科哪些是人工智能的基础：

经济学、心理学、数学、哲学

- 理性指的是一个系统的属性，即在已知的环境下做正确的事

- 对于如下比拟，哪些方法更通用并且适合科学开发？

理性动作、理性思考

- 理性智能体指的是有正确行为的智能体，但正确的行为意味着什么？

如果该序列是所期望的，则该智能体表现良好

智能体在某个环境中依据感知生成动作序列

这些动作序列引起环境改变而产生状态序列

- PEAS 是一种任务环境描述，它代表性能、环境、动作器、感受器

- 下列中语句哪些与智能体环境类型有关

阵发性与连续性

确定性与连续性

单智能体与多智能体

- 智能体函数与智能体程序的差异是：

一个智能体程序实现一个智能体函数

智能体程序包含智能体函数

- 选择下列哪些方法可以用于表示智能体的状态

因子式、结构式、原子式

- 凭借符号及他们之间（by南大ai院小妖兽）的关系来表征信息的人工智能方法是符号主义 AI
- 智能体可看作是通过感受器感知外部环境，并且通过执行器作用于外部环境的事物

by南大ai院小妖兽

百面机器学习

绪论

人工智能：使机器像人一样 A.感知 B.思考 C.做事 D.解决问题

机器学习的步骤：训练和测试，且可以重叠进行

模型：训练学习的结果

有监督模型：有老师告诉正确答案

无监督模型：仅仅靠观察自学

人工智能之父：麦卡锡、马文·明斯基

人工智能三次浪潮

1. 基于逻辑，上世纪 80 年代初因缺乏应用而进入“冬季”
2. 专家系统：人工智能与数据结合，把世界上所有的信息数字化。
3. 基于两个领域的发展：强大的计算能力，海量的数据

第三次人工智能浪潮和前两次的不同之处在于，人工智能离开了学术实验室，走进公众视野，应用更加普遍，对普通人的生活产生了影响。

特征工程

线性归一化

对原始数据进行线性变换，使之等比例映射到 $[0,1]$ 区间。

$$X' = \frac{X - X_{\max}}{X_{\max} - X_{\min}} \quad (\text{等比例缩放})$$

零均值归一化

将原始数据映射到均值为 0，标准差为 1 的分布上

$$z = \frac{x - \mu}{\sigma}$$

注：通过梯度下降法求解的算法通常需要归一化，但对于决策树模型并不适用

类别型特征

- 序号编码：类别间具有大小关系
- 独热编码：不具有大小关系，如（ABO，AB）血型，可变成一个 4 维向量
- 二进制编码：先用序号编码给每个类别赋予一个类别的 ID，然后根据 ID 对应的二进制编码作为结果。

模型评估：

准确率（accuracy）

- 分类正确的样本占总样本的比例
- 缺点：当不同类别的样本的比例极不平衡时，占比大的类别往往成为影响准确率的最主要的因素

精确率：分类正确的正样本，占经分类器判定的正比例

召回率：分类正确的正样本，占真正的正样本的比例

模型评估的方法：

Holdout 检验

- 将原始数据随机分为训练集和测试集
- 缺点：测试集计算出来的最后评估指标与原始分组密切相关

交叉检验

- ① 将全部样本划分为 k 个大小相等的样本子集
 - ② 依次遍历每个子集，当前子集作为测试集，其余 $k-1$ 个子集作为训练集
 - ③ 把 k 次评估指标的平均值作为最终指标
- 缺点：时间开销大

自助法

- 样本总数为 n
- 做 n 次有放回的随机抽样，得到大小为 n 的训练集
- 未抽到的样本作为测试集
- 当 n 很大时大约有 $1/e$ 比例 ($\approx 36.8\%$) 的样本从未抽到过，可以作为测试集

过拟合

- 模型在训练集表现好，在测试集或者加入新数据之后表现不好
- 多项式的阶数越高，拟合数据表现越好，但阶数太高会导致过拟合
- 处理方法：
 - ① 减小噪声
 - ② 降低模型复杂度，没用的特征就不要了
 - ③ 正则化：加入惩罚项
 - ④ 集成学习，如 Bagging

欠拟合

- 模型在训练集和预测时表现都不好
- 处理方法：
 - ① 添加新特征，新数据
 - ② 让模型更复杂，减小正则化系数

k 均值聚类

- 在事先不知道任何样本类别标签的前提下，通过数据之间的内在关系把样本划分为若干 类别，使得同类别样本之间的相似度高，不同类别样本之间的相似度低

- ① 预处理数据
- ② 随机选取 K 个中心点，记为 $\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_K^{(0)}$ 。
- ③ 定义代价函数（见书，先分类，再找中心点）

对每个点，找到离它自己最近的中心点，分成 K 类数据

对每类数据，重新计算中心，每点到中心的距离的平方和最小

- 缺点：
 - ① 受到初值和离群点的影响，每次的结果不稳定
 - ② 得到的通常是局部最优，不是全局最优
 - ③ 如果不同类型的数据的数量差别很大，很难有效的解决（1 类 1 个数据，1 类 10000 个数据）

- ④ 需要人工预先指定 K 值，和真实值未必吻合
- ⑤ 容易受到噪声影响
- ⑥ 样本只能被划分到单一的类型中

高斯混合模型

- 假设每个类型的数据都服从正态分布（高斯分布）

- 优点：

- ① 可给出一个样本属于某类数据的概率
- ② 可用于概率密度的估计
- ③ 用于生成新的样本点

学习中的知识

假反例：把实际是对的判断成了错的

假正例：把实际是错的判断成了对的

泛化：让假说包含假反例（初始条件太严格，以至于正确的数据不符合该条件，导致将其判断成了错误的）

狭化：让假说排除假正例（初始的条件太易满足，以至于错误的的数据符合该条件，判断为正确的）

假说 C1 是假说 C2 的泛化，则 C1 的条件没有 C2 严格，那么满足 C2 条件的数据肯定满足 C1

$\forall x C_2(x) \Rightarrow C_1(x)$ ，这叫做丢弃条件

将其他条件加入原始假说中，使其条件变得苛刻，或者删除原始条件中的析取式（A 或 B 称为析取范式，AB 条件满足一个即可，把他修改为两个条件，满足 A，且满足 B），可以狭化一个假说。

最大泛化假说：最宽松的假说，但与当前的观察保持一致（没宽松到犯错）

最大狭化假说：最严格的假说，但与当前的观察保持一致（没严格到犯错）

基于解释的学习（EBL）

把背景知识当成假设，比如看到有人用木棍烤东西吃，就假设所有木棍都能烤东西吃

对背景知识的解释充分，但没有从阳历中学到知识

基于相关性的学习 (RBL)

由背景知识和样例推理出一个新的假设，比如听到某给巴西人说西班牙语，一般常识是属于一个国家的人都说同一种语言，于是假设所有巴西人都说西班牙语，利用背景知识和看到的现象推理出假设，（背景知识 + 现象 \rightarrow 假设）但是该模型不能从零开始生成新知识。类比形而上学认识论，从实践得到认识，但认识却没有去指导新的实践

基于知识的归纳学习 (KBIL)

用背景知识和假设来解释样例，比如爱因斯坦基于黎曼几何（背景知识）提出广义相对论（新假设）来解释引力透镜（样例），在之前的知识的基础上大胆提出假设，来解释发生的现象！

现象 \rightarrow 背景知识、假设 \rightarrow 用假设解释现象，类比马克思主义认识论，得到的认识是为了解释现象，指导实践，看似是原点的回归，实际上是质的飞跃！

不确定性推理

- 推理分为经典逻辑推理和不确定性推理，不确定性推理的方法有：可信度方法、证据理论、模糊推理

- 产生式表示法又称为产生式规则表示法，运行于产生式系统
- 确定性规则知识的产生式： $P \rightarrow Q$ 或 (if p then q)
- 不确定性规则知识的产生式： $P \rightarrow Q(\text{置信度})$ 或 (If p then q(置信度))

例如，在疾病诊断专家系统中一条产生式：

if 本微生物的染色斑是革兰氏阴性 AND 本微生物的形状呈杆状 AND 病人是中间宿主 AND then 该微生物是绿农杆菌，置信度为 0.6

它表示当前条件都被满足时，结论的可相信的程度为 0.6

- 确定性事实性知识的产生式用三元组（对象，属性，值）表示

如，小明今年 18 岁：(Ming, age, 18)

- 不确定事实性知识的产生式用四元组（对象，属性，值，置信度）表示

如，李雷和韩梅梅不太可能是朋友：(friend, LL, HMM, 0.1)

- 蕴含式与产生式的异同：

蕴含式只能表示确定的知识，产生式可以表示确定或不确定的知识

蕴含式匹配必须是精确的，产生式不要求精确匹配，只要相似度落在指定范围内就认为他是可匹配的

除逻辑蕴含外，产生式还包括各种操作、规则、变换、算子、函数等。

例如，“如果炉温超过上限，则立即关闭风门”是一个产生式，但不是蕴含式。

- 产生式的形式描述及语义——巴科斯范式 BNF

<产生式> ::= <前提> -> <结论>

<前提> ::= <简单条件> | <复合条件>

<复合条件> ::= <简单条件> AND <简单条件> [AND <简单条件> ... | <简单条件> OR <简单条件> [OR <简单条件> ...]

<结论> ::= <事实> | <操作>

<操作> ::= <操作名> [(<变元> , ...)]

符号 "< >" 表示 "必选项" ; 符号 "[]" 表示 "可选项" 。符号 "::=" 表示 "定义为" ;

符号 "|" 表示 "或者是"

产生式系统

- 把一组产生式放在一起，让他们互相配合，协同作用，一个产生式生成的结论可以供另一个产生式作为已知事实使用，以求得问题的解决，这样的系统 称为产生式系统

产生式系统由三个部分组成：控制系统、规则库、综合数据库

1. 规则库：

- 啥是规则？ if ... then... 或 $P \rightarrow Q$
- 规则库又是啥？ 用于描述相应领域中的知识的产生式集合称为规则库，

$\{P_i \rightarrow Q_i\}$

- 规则库有啥用？ 知识是否完整，一致，表达是否准确灵活，对知识的组织是否合理，直接影响到系统的性能，所以一个好的规则库：知识完整，知识一致，表达灵活，组织合理

- 建立规则库时，应注意以下问题：

有效的表达领域内的过程性知识

对知识进行合理的组织与管理，排除那些冗余和矛盾的知识

- 冗余：不同的人对同一个规则库进行完善的时候，可能会出现重复，同一种事物的判断，条件可能相差不大

- 矛盾：由于对事物的认知程度的不同导致统一现象有着不同的前件，典型的例子（盲人摸象）

2. 综合数据库——存放问题求解过程中各种当前信息的数据结构

- 当规则库中的某条产生式的前提可与综合数据库中的某些已知事实匹配时，该产生式就被激活，并把用它推出的结论放入综合数据库中，作为后面推理的已知事实
- 综合数据库的内容是在不断变化的，是动态的。
- 综合数据库中的已知事实通常用字符串、向量、集合、矩阵、表等数据结构表示。

3. 控制系统——又称为推理机构，由一组程序组成，负责整个产生式系统的运行,实现 对问题的求解

· 通俗的讲，控制系统就是怎样去组织知识、证据、事实等信息，并最终获取一个 确定的结论的过程

· 它主要作以下几项工作:

1. 按一定策略从规则库中选择规则与综合数据库中的已知事实进行匹配。
2. 匹配成功的规则可能不止一条，这称为发生了冲突。此时，推理机构必须调用相应的解决冲突的策略进行消解
3. 在执行某一条规则时，如果该规则的右部是一个或多个结论，则把这些结论加入到综合数据库中; 如果该规则的右部是一个或多个操作，则执行这些操作。
4. 对于不确定性知识，在执行每一条规则时还要按照一定算法计算结论的不确定性。
5. 随时掌握结束产生式系统运行的时机，以便在适当的时候停止系统的运行。

· 举例说明产生式系统的求解过程:

例: 设在综合数据库中存放有下列已知事实:

该动物身上有暗斑点，长脖子，长腿，有奶，有蹄，且假设综合数据库中的已知事实与规则库中的知识是从第一条开始，逐条进行匹配的，则推理机构的工作过程如下:

1. 从规则库中取出第一条规则 r1，检查前提条件与综合数据库中的已知事实不匹配; 取第二条规则 r2，r2 的前提“该动物有奶”与综合数据库中事实匹配，则 r2 被执行，其结论被加入综合数据库中，此时综合数据库中的事实为:

该动物身上有暗斑点，长脖子，长腿，有奶，有蹄，是哺乳动物

2. 接着取 r3 r4 r5 r6 都不匹配，取到 r7 时，匹配成功，则将 r7 的结论加入综合数据库:

该动物身上有暗斑点,长脖子,长腿,有奶,有蹄,是哺乳动物,是有蹄动物

3. 接着取规则，取到 r11 时，匹配成功，发现其前提条件与综合数据库完全匹配，

则确定该动物是:长颈鹿

至此，问题的求解结束了。

暗斑点，长脖子，长腿，有奶，有蹄					
	IF				THEN
1	毛发				哺乳动物
2	有奶				哺乳动物
3	羽毛				鸟
4	会飞	下蛋			鸟
5	吃肉				肉食动物
6	犬齿	有爪	眼盯前方		肉食动物
7	哺乳动物	有蹄			蹄类动物
8	哺乳动物	嚼刍动物			蹄类动物
9	哺乳动物	肉食动物	黄褐色	暗斑点	金钱豹
10	哺乳动物	肉食动物	黄褐色	黑色条纹	老虎
11	蹄类动物	长脖子	长腿	暗斑点	长颈鹿
12	蹄类动物	黑色条纹			斑马
13	鸟	长脖子	长腿	黑白二色	鸵鸟
14	鸟	游泳	不飞	黑白二色	企鹅
15	鸟	善飞			信天翁

· 以上介绍了不确定推理的基本概念，以下为三种不确定推理方法：

可信度方法

· 可信度是啥？根据经验对一个事物或现象为真的相信程度称为可信度

例如:小李今日上班迟到了，理由是路上自行车出了毛病？

从小李自己这方面来看，“自行车坏了”这个理由有两个情况，真和假。

从听众的角度来分析，只是有某种程度的确信，而相信的程度是依据小李的平时表现。

所以可信度的概念是: 经验+现象+相信程度

· 可信度带有较大的主观性和经验性，其准确性难以把握。但人工智能所面对的多数是结构不良的复杂问题，难以给出精确的数学模型，先验概率及条件概率的确定也比较困难，因而用可信度来表示知识及证据的不确定性仍不失为一种可行的方法。

· CF 模型是基于可信度表示的不确定性推理的基本方法。其它可信度方法都是在此基础上发展起来的。

- 知识不确定表示: IF E THEN H 或写作 $CF(H,E)$

- $CF(H,E)$ 是该条知识的可信度, 称为可信度因子或规则强度, 也就是前面所说的静态强度。可信度因子用来评价证据的不确定性

- $CF(H,E)$ 在 $[-1, 1]$ 上取值, 它指出当前提条件 E 所对应的证据为真时, 它对结论为真的支持程度。

例如: if 头痛 and 流涕 then 感冒 (0.7)

表示当病人确有“头痛”及“流涕”症状时, 则有 7 成的把握认为他患了感冒。

任何一条知识都不是独立存在, 每条知识都存在都依赖于外部环境。

- 对于初始证据, 其可信度的值由提供证据的用户给出。

- 对于用先前推出的结论作为当前推理的证据, 其可信度值通过不确定性传递算法计算得到。

- $CF(H, E)$ 反映了前提条件和结论的联系强度, 是相应知识的知识强度

- 因此, $CF(H,E)$ 的值要求领域专家直接给出。其原则是:

- 若由于相应证据的出现, 增加了结论 H 为真的可信度, 则使 $CF(H,E) > 0$, 证据的出现越是支持 H 为真, 就使 $CF(H,E)$ 的值越大

- 反之, 使 $CF(H,E) < 0$, 证据的出现越是支持 H 为假, 就使 $CF(H,E)$ 的值越小

- 若证据的出现与否与 H 无关, 则使 $CF(H,E) = 0$

组合证据

- 当组合证据是多个单一证据的合取时, 即: $E_1 \text{ and } E_2 \text{ and } \dots \text{ and } E_n$, 若已知 $CF(E_1)$, $CF(E_2)$, ..., $CF(E_n)$, 则 $CF(E) = \min \{CF(E_1), CF(E_2), \dots, CF(E_n)\}$

- 当组合证据是多个单一证据的析取时, 即: $E = E_1 \text{ or } E_2 \text{ or } \dots \text{ or } E_n$ 若已知 $CF(E_1)$, $CF(E_2)$, ..., $CF(E_n)$, 则 $CF(E) = \max \{CF(E_1), CF(E_2), \dots, CF(E_n)\}$

- CF 模型中的不确定性推理是从不确定的初始证据出发，通过运用相关的不确定性知识，最终推出结论并求出结论的可信度值。结论 H 的可信度由下式计算：

$$CF(H) = CF(H,E) \times \max \{ 0, CF(E) \}$$

带有阈值限度的不确定推理

知识的表示形式：

知识表示为: if E then H (CF(H, E), λ)

其中 λ 是阈值，它对相应知识的可应用性规定了一个限度: $0 < \lambda < 1$

1. E 为知识的前提条件，H 为结论
2. CF(H,E)为知识的可信度，即规则强度，相应知识为真的可信程度[0,1]
3. λ 是阈值，当前提条件 E 的可信度超过这个限度，相应的知识才可能被应用

不确定性的传递算法

$$CF(H) = CF(H,E) \times CF(E)$$

结论不确定性的合成算法

设有多条规则具有相同的结论，即

IF E1 THEN H (CF(H,E1), λ_1)

IF E2 THEN H (CF(H,E2), λ_2)

...

IF En THEN H (CF(H,En), λ_n)

且满足 $CF(E_i) \geq \lambda_i$

证据理论

Ppt 6.3