AI复习

第2章 逻辑与推理

- 命题逻辑
- 谓词逻辑
- 知识图谱概述
  - 知识图谱定义
    - 有向图，描述实体和实体之间的关系
    - 一阶逻辑CNL三元组
- 知识图谱推理
  - 归纳逻辑程序设计ILP
    - FOIL归纳推理 p54
      First order inductive learner
      - 序列覆盖
        思路：一条规则特殊，逐步添加目标谓词的新谓词来修正，直到所有或的推理规则不覆盖任何反例
      - 输入：目标谓词、训练样例（正例反例）、其他背景知识规则
      - 输出：可推到目标谓词或它的推理规则
      - 算法：
        ① 将目标谓词作为结论
        ② 加进原问作为前提的最谓词，这一条加到推理规则中，计算FOIL信息增益函数
        ③ 选可以带来最大信息增益的谓词并来谓词加入推理规则，移出新的推理规则
        ④ 剪除训练样例覆盖中生成推理规则不符的样例
        ⑤ 重复2-4直到推理规则不覆盖任何反例
  - 推理方法
    - 路径排序算法PRA
      - 思路：将实体之间的关联路径作为特征，学习路径的权重的分类器
      - 算法：
        ① 特征编取：生成并选择路径特征集合
        ② 特征计算：计算每个训练样例的特征值P(s→t,f)，可以是概率、bool、频次等
        ③ 分类器训练
    - 基于分布式表达的知识推理

第3章 搜索算法

- 3.1 搜索算法基础
  - 基本定义
  - 评价指标
    - 完备性：存在解，算法能否找到
  - 算法框架
    - 边缘集合-开表
      - 扩展
        - DFS
        - BFS
        - 无信息搜索
    - 剪枝
  - 图搜索和树搜索
    - 同/不同扩展边缘结点
    - 优势：储存所有被扩展过的结点的状态
    - 图搜索扩展的结点更少（因为剪枝，优势）
    - 但可能找不到最短路径
    - 树只重复搜索上一层的多条路径的结果
- 3.2 启发式搜索（有信息搜索）
  - 基本概念
    - 启发函数h(n)
      - 估计结点n距离目标结点还剩多少代价
      - 常用来指定pop_from函数
    - Heuristic(启发式)搜索h(n)
      - 决定搜索算法扩展结点n的优先度
    - 评价函数f(n)
      - f(n)部分，被挑选优先级更高
  - 贪婪最佳优先算法
    - f(n) = h(n)
    - 每次选择剩余路径的最优方法！
    - 问题：但可能找不到最优解
  - A*算法
    - g(n)：起始点 - 结点n的最佳代价
    - 定义f(n) = g(n) + h(n)
    - （不重求）一致代价搜索UCS
      - f(n) = g(n)
    - 特征：信息确定、全局可观察、竞争对手轮流行动、输赢收益零和
    - 定义棋局第n回合末决策：
      在第n+M次棋盘大状态中的局势分；
      在第n+M回合的局势分；
      - 在计算棋局时神每个结点分数后，每一步选择最合理的走法直接使用Minimax or αβ剪枝法，四个回合
      - MAX玩家在下一回合MAX玩家给出的选择是我最大的；
      - MIN玩家在下一回合MAX玩家给出的选择是我最小的；
      - 递归下去直到叶子结点(terminal)，在root返回
      - 缺点：每次要重新开到最底下，复杂度很高
      - 常规关键：算法对每个结点有一个明确的收益估值
      - 思路总结：p47
      - 我们有一个定义f(s→s)的估值函数，结点态最佳上面的MAX结点在我的收益为alpha，
      - 一个定义β(s→s)的估值函数，是为s的估计或估计一部分后，节数s最佳上面MAX结点目前的最小alpha，而s的未来决策的结果会并被剪枝。
    - Alpha-Beta剪枝算法
      - 初始化为(-∞, +∞)
      - 根多父结点的alpha和beta
      - 更新
        - 对子Max结点：若孩子中的的收益>alpha，更新alpha
        - 对子Min结点：若孩子MAX的收益<beta，更新beta
      - 剪枝
        - 孩子父结点下面
        - 孩子父结点上面
      - 若alpha>beta，该结点后续未访问结点被剪枝
- 3.3 对抗搜索（博弈搜索）
  - 最大最小搜索 minimax search
  - 为减少时间开销 剪枝技术
- 3.4 强化学习
  - 问题：不能预先计算最优策略，需要一边探索一边决策
  - 目标
    - 最大化奖励
  - 贪心算法
    - 探索
      - 尝试新的选项
    - 利用
      - 根据已知信息做出最佳决策
    - epsilon - 贪心算法
      - 有epsilon概率随机选
      - 没有考虑强的次数，不合理
      - 之前没有以大概率选一个机器的数据分布了
  - 上置置信区间算法 UCB1
    upper confidence bounds
    - 为每个动作的奖励期望计算一个估计范围，优先未来给介于最上限的动作
    - 如何计算置信期望的估值范围
      - 霍夫丁不等式
      - pPS 式5.5
    - 策略：默认选择动作a_t，使（式5.5）上限最大
  - 蒙特卡洛树搜索
    - 要求降低，只找近似最优解
    - 算树方法
      - 1. 选择
        - UCB1
        - 从root开始向下递归选择子结点，直到leaf或者具有未扩展children的结点。
      - 2.扩展
        - 如果L不是terminal，随机扩展一个它的未曾扩展的child M
      - 3. 模拟
        - 从M出发，随机走到终局，即走到leaf
      - 4.反向传播
        - 用终局得分回溯更新沿模拟路径中M到root之间的路（包括M）的统计量（被访问次数）
    - 上置置信区间树搜索 UCT
      - 基于最大最小算法
      - 用UCB1
    - 蒙特卡洛采样
    - 详细举例：看p30题即可

第4章 机器学习-监督学习

- 目标：从数据集提取特征，学习一个映射函数f，用特征映射到最大空间，寻找数据特征间的关系
- 基本概念
  - 监督学习：无监督学习、强化学习区别
    - 监督学习：
      - 有标签信息
      - 目标：分类、回归
      - 一般包括三部分
        - 学习映射函数f
        - 测试集上测试f
        - 未知数据集上测试f（真正投入使用）
    - 无监督学习：
      - 无标签
      - 目标：聚类、降维
    - 半监督学习：
      - 部分标注数据
    - 强化学习：
      - 序列决策/决策学习，从交互中学习
    - 分类和回归的区别
      - 分类输出：有文空量
      - 回归输出：真实数值
  - 常用损失函数
    - 0-1损失函数
      - 分类任务
    - 平方损失函数
    - 绝对损失函数
    - 对数损失函数/对数似然函数
  - 经验风险和期望风险
    - 经验风险小-拟合好
      期望风险小-模型好
    - 训练集loss最小-是否经验风险最小化
      机器学习目标：期望风险最小化
    - 期望风险是模型关于联合分布的期望损失，经验风险是关于训练样本集的平均损失
  - 结构风险最小化
    - 平衡模型复杂度和最小化经验风险
      - 正则化项
      - 惩罚项
    - 主要的监督学习方法
      - 判别方法
        - 直接学习判别函数f(x)或条件概率分布P(Y|X)作为模型
        - 回归模型、神经网络、支持向量机、Ada boosting
      - 生成方法
        - 学习联合概率分布P(X, Y)
        - 特点：联合分布概率或似然概率模型
      - 最佳：mean/残差平方最小
      - 参数求解过程：最小二乘法
  - 4.2 回归分析
    - 一元线性回归
    - 多元线性回归
    - 参数求解：最大似然估计法
    - 逻辑斯蒂回归(分类)/多项回归
      - 拟合曲线概率定义（a为正例的概率），常用于二分类问题
      - 输出具有概率定义（a为正例的概率），常用于二分类问题
      - 几率
        - 正概率/负概率
      - 对数几率/Logit/几率
        - log几率
      - 对输入数据x作为正例的概率取对数，就得到了线性回归模型log(P(Y=1|x)/(1-P(Y=1|x)))=w·x
      - 实质：用线性回归模型的输出，拟合真实标记的对数几率
      - 推导(看上面的概率-负概率，即P(y=1|x)-0)=0，解出下Logit公式，在原基w·x+b上求解使得目标最大值（线性模型，可以用线性函数w·x+b表示+Sigmoid函数）
  - 4.3 决策树
    - 信息熵
    - 信息增益
    - 信息增益率
    - Gini系数
  - 4.4 线性判别分析LDA
    Fisher线性判别分析FDA
    - 基于监督的降维方法
    - 类间散度、和Label
      最佳Label的目标是使投影后信息增益可以最大，不同类样本尽可能远离
      - 这个嵌入的线性变换，存在一个特别的 类别分布，即同类别数据可以被映射到一起
    - 变量计算步骤
  - 4.5 Ada Boosting
    - 集成若干个弱分类器，完成复杂的分类任务
    - 引入概念：大概近似正确 PAC
    - 核心问题
      - 如何提高被错误分类的样本的 权重
      - 如何提高分类错误小的分类器的权重
    - 步骤

第5章 统计机器学习-无监督学习

- 5.1 聚类算法
  - 目标：将数据集分成多个簇，使得内方差最小化
  - 要求：特征变量连续，没有异常值
  - 目标：将n个维数据划分为k个聚簇，使得内方差最小化
  - 问题
    - 高维点对均值影响大
    - 使用其他指标函数，或者K-medoids（选择聚簇中的真实点，取代质心）
    - 聚类算法聚类数目k
    - 对数据可视化做法，各个维度单位量纲一
    - 硬分类
      - 软分类：高斯混合模型
    - 现代算法
      - 层次聚类
      - 谱聚类——基于图谱的聚类算法
    - 应用
      - 图像压缩
      - 文本聚类
  - 5.2 降维
    - 其他用途
      - 代替PCA
      - 重构特变换
      - 条证正交分解PCO
    - 要求：降维后的结果保持原始数据的原有结构，数学角度来说就是保持原始高维数据的总体方差结构
    - 方差、协方差和相关系数
    - 步骤
      - 对每个记录进行中心化处理
      - 求对协方差矩阵
      - 求协方差矩阵的特征值分解，对特征向量排序
      - 选择前个最大特征向量对应的特征向量组成转换矩阵W
    - 线性降维
    - 非负矩阵分解NMF
      - 非负大矩阵分解成非负小矩阵
      - 应用：图像像素矩阵、自然语言语言-文档矩阵
      - 要求分解后所有元素均为非负值
      - 非线性降维
    - 非线性降维方法
      - 保持原始数据两两距离不变
      - 多维尺度法MDS
      - 根据距离矩阵计算，保证降维后数据距离关系保持不变
      - 无法对新数据集合进行降维，因为没有距离矩阵
      - 线性降维
      - 局部线性嵌入LLE
      - 基本假设：
        一个流形的局部可以近似于一个欧式空间，每个点可以用它的邻居线性表示。
        降维后，原来高维空间中的邻居关系，在低维空间中仍然成立。通过局部重建权值矩阵来保持局部结构信息，最终保持数据整体几何信息。
      - 非线性降维
  - 5.3 特征人脸方法
    - 应用PCA实现人脸图像降维
    - 原始数据X是n个1024维的人脸向量，进一步PCA，只为得出个特征人脸（特征向量），表达，每个特征向量都是1024维
    - 即X：n*1024，W：1024*36
  - 5.4 潜在语义分析
    - 奇异值分解，重建
  - 5.5 期望最大化算法
    - E 求加权量
    - M 期望最大化
    - 不能保证找到全局最大值，有可能是鞍点
    - 二维半三硬币

第6章 深度学习

- 6.1 发展历史
- 6.2 前馈神经网络 feedforward neural network
  - 概念
    - 神经元
      - 为了能够使用梯度下降，激活函数必须连续可导
      - sigmoid
      - 激活函数
      - tanh
      - ReLU
      - 修正线性单元
      - softmax
      - 多分类，求输入数据出现在类别上的概率分布
    - 损失函数
      - 均方误差函数MSE
      - 交叉熵损失函数
      - 可用来度量两个概率分布之间的差异
    - 梯度下降
      - 批量梯度下降
      - 随机梯度下降
      - 小批量梯度下降
    - 反向传播
    - 梯度消失问题
    - 多层感知机
  - 参数优化和学习
    - 梯度下降
      - 参数求导
      - 输出层
    - 结构
      - 输入层
      - 隐藏层
      - 输出层
      - 同一层神经元不相互连接，不同层神经元只有相邻的连接
  - 6.3 卷积神经网络
    - 空间卷积
    - 卷积核
      - 用有卷积kernel
      - 高斯卷积核
      - 中心最大，向四个角的过程中逐渐减少
      - 图像平滑操作：
        当卷积核中心权重较小，相其他卷积核的重新赋值到最小，同时保持图像边缘清晰度
      - 梯度锐化，卷积内容，减少卷积防止过拟合
    - 池化层(Pooling)
      - 防止存在下采样
      - Binding
      - 卷积滤波结果图
      - 特征图
      - 信息压缩
      - 下采样操作
    - 卷积层
      - 每个输出的特征图由数个，输入图像卷积后的特征图加成后的结果
      - 卷积层
      - 滑动窗口——下采样操作
      - 滑动窗口上一个点对应输入图像上的区域
    - 优化
      - 对特征图进行代表下采样
      - 最大池化
      - 平均池化
      - K最大池化
      - 动态K池化
      - 全局平均池化GAP
      global average pooling
      - 分类式卷积层
      - 初步卷积学习
      - 卷积，激活函数-池化-全连接层
    - 卷积神经网络
      - 提升泛化能力，减少过拟合
      - 随机丢掉一部分神经元
      - Dropout
      - 每次迭代训练中，以一定概率随机剔除每一层中部千神经元
      - 卷积
      - 卷积神经网络中每一层任意神经元的输入是与该卷积核有关的区域，即Receptive field(感受野)的元，强制保持卷积结构基本不变分布
      - 卷积
      - 对表示函数的 input卷积核的特征图和大区域，是局部概念关系问题，提高卷积速度
      - 卷积
      - 卷积层的LTM
      - 卷积层的LTM
      - 卷积层的LTM