

5/12 周记

一，继续学习python

本周学习：

- > 06 第六章python 数据容器
 - > 07 第七章python函数进阶
 - > 08 第八章python 文件操作
 - > 09 第九章python异常 模块 与包
 - > 10 第十章 基础综合案例
 - > 11 第十一章 地图可视化
 - > 12 第十二章 动态柱状图
 - > my_package
 - > my_package1
 - > 可视化案例数据
- > 2-01 面向对象

二，外文文献阅读情况

共三篇 均为关于 大语模型结合实际工程示例

[illegible]

文章信息	背景、目的与结论	结果与讨论	文章好在 哪里	自我想法
	fields. 结论: In conclusion, this study demonstrates that a fine-tuned data indexed GPT model can significantly improve query response performance compared to state-of-the-art GPT-4.			

三，博士论文阅读第2、3 章

第二章 建模表示与建模方法

1，为什么加工工艺过程信息集成形成的数据流是四个，而且系统集成接口是什么

答：根据产业4个核心的数据要求划分，规范化制造设备、业务流程管理、制造过程数据采集、储存

答：四个中心的数据转换接口，如物理信息的转换，plc接口输出的信息转化为json格式，化学液体信息的转 换

2，系统建模是什么

答：类比于施工图，功能、参数用标准化图纸画出来，让所有参与者均可根据该图纸进行项目设计。利用 IDEF,UML两个图形化工具，产出一个模型文档

IDEF : 描述系统功能和数据的关系 结构化建模 分析阶段

IDEF0：功能建模 分解系统功能，用数据流写出系统输入输出关系，但不对系统动态行为进行描述也不支持 具体功能的实现

IDEF1x：数据建模 描述系统信息与关系，但面向结构化数据，无法表示操作，对面向对象支持不足

UML: 全过程设计 面向对象建模

用例图：从用户角度描述系统功能与交互 可以直观显示用户需求和系统功能的对应关系，但仅设计了对外可 见的功能， 不涉及内部具体实现，并且要按照特定的设计模式设计系统

类图：结构建模 对系统内部对象的进行设计 可以精确描述对象的结构与行为，直接指导代码开发，适合复杂 系统模块化设计（和最近学的python也提到这一点）

二者结合满足了滚磨光整加工数据库构建的要求

3，为什么要从五个方面进行系统建模 功能 组织 信息 知识 过程 其中过程是其余的集成

答：一个复杂流程的问题，分析为几个小问题解决

3.1，问功能有什么，如何做？遇到的数据如何保存？具体工艺如何选择？如何串联起来形成一个大系统，他们各自用了什么建模方法

功能模型：IDEF0结构化建模 功能树图

组织模型中：**ARIS-Toolset组织模型** 一中组织模型结构有四个对象 不用uml idf 因为不是专业的企业架构，缺少单元项目组的关系概念

信息模型中：模式分解 IDEF1x+规范化理论 严格遵守数据库规范，设计数据库表结构

知识模型：知识语义网络 支持案例推理，专家经验建模

过程模型：由逻辑关系进行推理

4，体系结构是什么 如何划分的

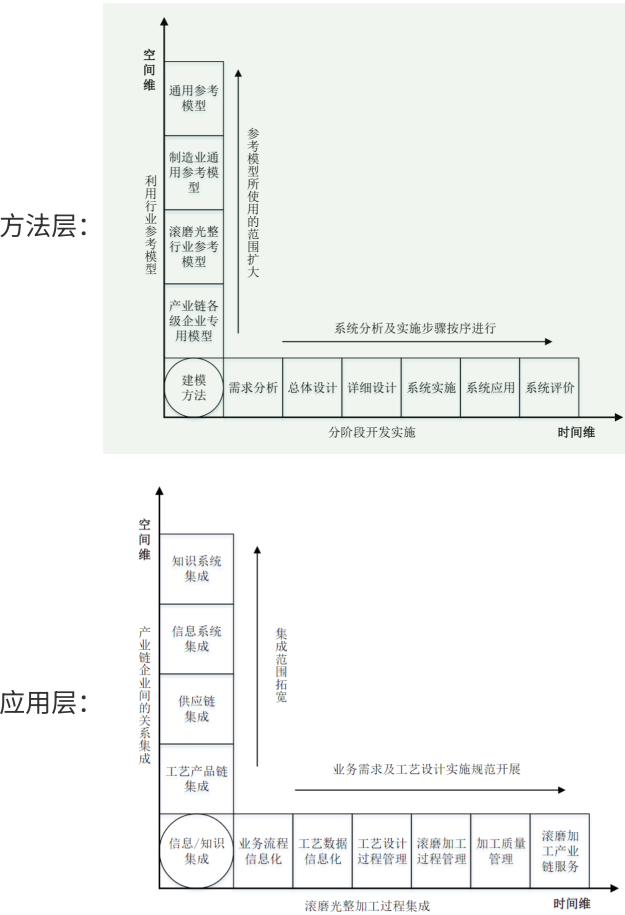
数据库体系结构，支撑全产业链信息集成与智能化的顶层框架

CIM-OSA:一个三维模型 生命周期维，视图维，通用维。

ARIS: 企业建模集成方法 功能，组织，数据，控制

L³体系结构: 视图层，方法层，应用层每层包含空间维，时间维

视图层：五大模型存放 按时间空间划分



应用层可以向智能化的方向研究，

第三章

现状：大量合格实例未利用 → 需求：构建案例库实现智能优选

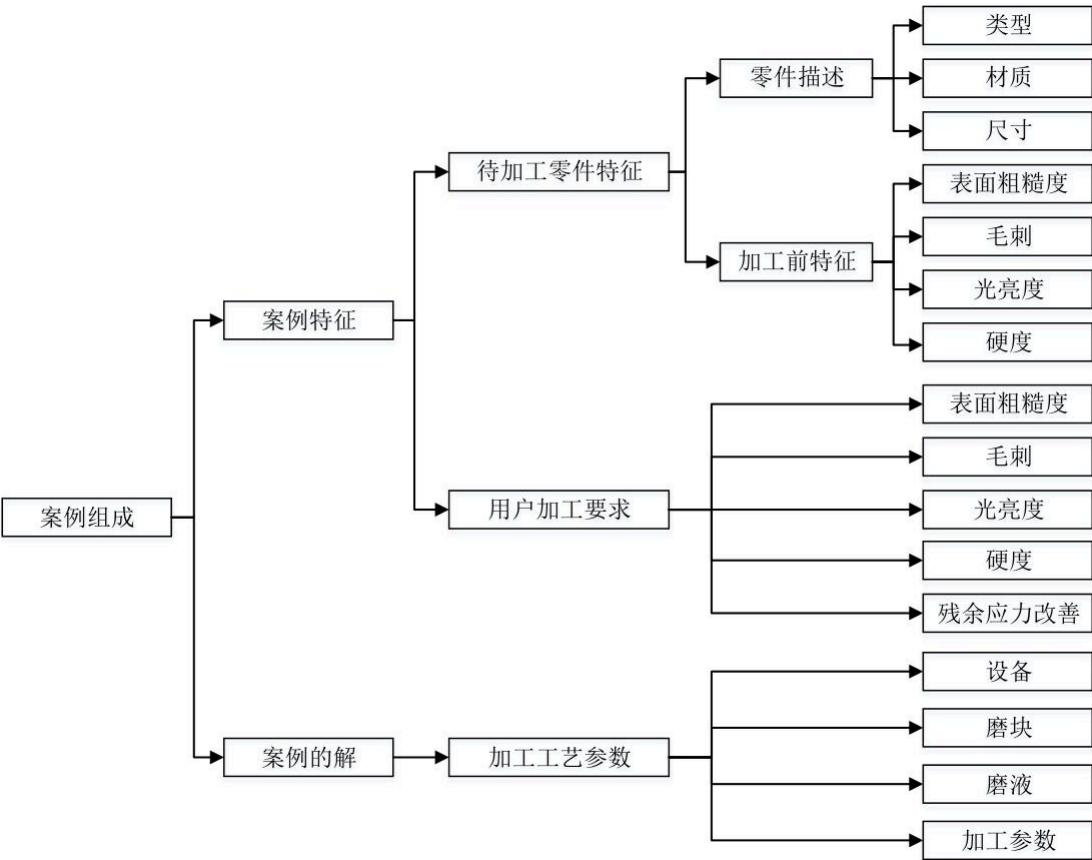
↓ 方法：CBR（快速匹配）+ FES（补充推理）融合推理

↓ 关键：案例库需合理构建与优化（解决冗余问题）

↓ 方案：S-FCM算法（减法聚类改进FCM，聚类后去冗余）

1，案例表征过程，如何确定的问题是什么，解空间是什么，问题空间是什么

答：相关参数由大量的实验报告与生产实例来确定，而问题就是零件的加工需求，需要提取案例特征来定义（待加工零件的特征，用户的加工要求）；问题空间是，案例特征的取值范围构成的集合，案例库需要覆盖问题空间的不同区域，避免出现无解区域；解空间就是可行工艺参数的集合，包括设备，模块，磨液，加工参数



2，案例库优化方法是什么，

答：已知：案例库中的所有案例均为合格案例，先利用FCM算法找到案例库中的孤立案例，保留这种案例， 然后进行相似度计算，删除冗余案例

3，FCM聚类算法基本原理与代码实现

FCM 算法

FCM聚类属于一种软聚类，两个参数，样本与聚类中心的距离、聚类中心的位置
然后构造一个损失函数：

$$\min(L) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2$$

μ 是一个加权效果，距离远，权重低，距离小，权重大,且有 $\sum_{j=1}^C \mu_{ij} = 1$ ； m 是超参数，是一个事先准备的参数 m 过大

时，相当于 $\min(L) = \sum_{i=1}^N \sum_{j=1}^C \|x_i - c_j\|^2$ 最小平方误差，加权消失； m 过小时，变成硬聚类；简而言之，求两个参数什么时候会使得损失函数最小。

约束条件：M个样本点，均需要满足 $\sum_{j=1}^C \mu_{ij} = 1$,

由此得：拉格朗日函数

$$\begin{aligned} L(\mu, c, \lambda) &= \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2 + \sum_{i=1}^N \left(\lambda_i \left(\sum_{j=1}^C \mu_{ij} - 1 \right) \right) \\ &= \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2 + \sum_{i=1}^N \sum_{j=1}^C (\lambda_i \mu_{ij} - \lambda_i) \end{aligned}$$

1 | 求导数即可，分别有：对 μ 求导

$$\begin{aligned} \frac{\partial L}{\partial \mu_{ij}} &= m \cdot \mu_{ij}^{m-1} \|x_i - c_j\|^2 + \lambda_i \\ \mu_{ij}^{m-1} &= -\frac{\lambda_i}{m \|x_i - c_j\|^2} \\ \Rightarrow \mu_{ij} &= \left(-\frac{\lambda_i}{m \|x_i - c_j\|^2} \right)^{\frac{1}{m-1}} \\ \Rightarrow \sum_{j=1}^C \left(-\frac{\lambda_i}{m \|x_i - c_j\|^2} \right)^{\frac{1}{m-1}} &= 1 \\ \Rightarrow \left(-\frac{\lambda_i}{m} \right)^{\frac{1}{m-1}} \sum_{j=1}^C \frac{1}{\|x_i - c_j\|^{\frac{2}{m-1}}} &= 1 \\ \Rightarrow \left(-\frac{\lambda_i}{m} \right)^{\frac{1}{m-1}} &= \frac{1}{\sum_{j=1}^C \frac{1}{\|x_i - c_j\|^{\frac{2}{m-1}}}} \\ \Rightarrow -\frac{\lambda_i}{m} &= \left(\frac{1}{\sum_{j=1}^C \frac{1}{\|x_i - c_j\|^{\frac{2}{m-1}}}} \right)^{m-1} \end{aligned}$$

回代求导结果 最终有: 实际上是一个概率 上面是一个 下面是整体, 最后正则化

$$= \frac{\left(\frac{1}{\|x_i - c_j\|} \right)^{\frac{1}{2(m-1)}}}{\sum_{j=1}^C \left(\frac{1}{\|x_i - c_j\|} \right)^{\frac{1}{2(m-1)}}} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{1}{2(m-1)}}}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

1 而对c求导有：c为一个矢量，对矢量求导时，需要对其分量分别求导然后拼接成一个矢量，推到结果也是一个概率形式，下面是属于c的所有，上面是一个加权x后的求和

$$\begin{aligned} \frac{\partial L}{\partial c_j} &= \frac{\partial}{\partial c_j} \left(\sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2 \right) + \frac{\partial}{\partial c_j} \left(\sum_{i=1}^N \lambda_i \left(\sum_{j=1}^C \mu_{ij} - 1 \right) \right) \\ &= -2 \sum_{i=1}^N \mu_{ij}^m (x_i - c_j) = 0 \\ \Rightarrow \sum_{i=1}^N \mu_{ij}^m x_i &= c_j \sum_{i=1}^N \mu_{ij}^m \\ \Rightarrow c_j &= \frac{\sum_{i=1}^N \mu_{ij}^m x_i}{\sum_{i=1}^N \mu_{ij}^m} \end{aligned}$$

综上所述,设定参数 μ, c, m_0 ,最大迭代次数与收敛精度,更新c和 μ 矩阵,重复代回,最后让 μ 矩阵满足收敛精度或者满足迭代次数(代码暂无)

减法聚类

由于SCF对初始聚类数和聚类中心敏感，避免陷入局部最优解时，应该利用减法聚类提供优质的聚类中心，核心思想是（豆包）：高密度区域的点优先被选为聚类中心，并通过“减法”操作抑制相邻区域的密度，避免邻近中心的重复选择。

原始密度公式 x_i 的密度指标定义：

$$D_i = \sum_{j=1}^n \exp \left(-\frac{\|x_i - x_j\|^2}{(\sigma R_a)^2} \right)$$

1 $\$R_a\$$ 可以通过分析数据点间的距离关系，先确定每个点到其他点的最大距离，再取这些最大距离中的最小值，最后取半，从而确定 ra 的具体值

$$r_a = \frac{1}{2} \min_j \left\{ \max_i \{ \|s_i - s_j\| \} \right\}$$

选择最大密度作为第一个聚类中心,然后归一化

$$D_c = \frac{D_c}{\max(D_1, D_2, \dots, D_n)}$$

1 | 密度衰减操，对其他xj的Dj进行一个修正 这里的r可以取1.2ra

$$D_j = D_j - D_c \cdot \exp \left(-\frac{\|x_j - c\|^2}{(\sigma R/2)^2} \right)$$

1 | 再这中选择最大的 作为下一个聚类中心,用以下方法判断何时停止

$$\max(D_i) < \delta \cdot D_{\text{first}} \quad \text{或} \quad \frac{D_{\text{new}}}{D_{\text{first}}} < \gamma$$

S-FCM算法

1,确定迭代次数,收敛精度,模糊指数m,2,离差归一化,消除量纲差异,3,减法聚类得到聚类中心V和聚类数c,进行FCM算法迭代5,得到最终的C和M后,求出不同聚类数下的聚类有效性函数

$$V_{\text{sie}}(R, V, c) = \frac{\frac{1}{M} \sum_{k=1}^c \sum_{i=1}^M r_{ki}^{m_0} \|v_k - s_i\|^2}{\min_{k \neq l} \|v_k - v_l\|^2}$$

不断减小c和v,依次取前c次聚类中心,最后将v_sie最小的输出其聚类数c,聚类中心V,隶属度矩阵R

1. 特殊案例筛选

- 定义**隶属度阈值**：若案例对所有类的隶属度均小于 u ，则视为**孤立案例（特殊案例）**，予以保留。

2. 冗余案例删除

- 计算同类案例的相似度阈值：若“案例对”相似度超过阈值，则删除其中一个案例（相似度公式 3-16，基于特征权重的加权平均）。
- 保留案例组成精简案例集 B2，与特殊案例集 B1 合并为新案例库

4, 如何进行仿真,或者如何取选择合适的δ,最后比对结果如何做?