**连续热镀锌机组班组操作经验提取 - 初步方案**

**东北大学 唐立新 教授团队**

**一、问题描述**

连续热镀锌机组生产的产品规格较多，关键控制参数众多，依据人工经验进行操作经常产生不同班组生产指标波动较大的情况，因此，拟通过班组的历史操作大数据，提取出不同规格下较好的操作经验，使得各班组之间的生产指标均达到较好范围。

**二、数据要求**

需要针对不同规格的产品，提供不同班组在生产时的实际工艺参数的设定值，以及所对应的评价指标：产量、产品质量。

数据量尽量大，以保证提取的结果能够反应实际的生产规律。

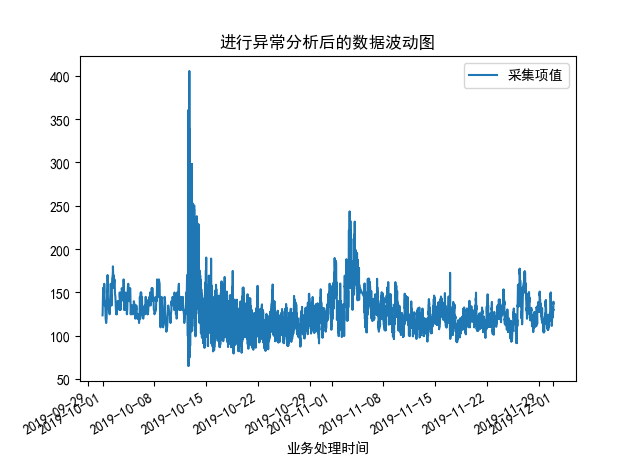
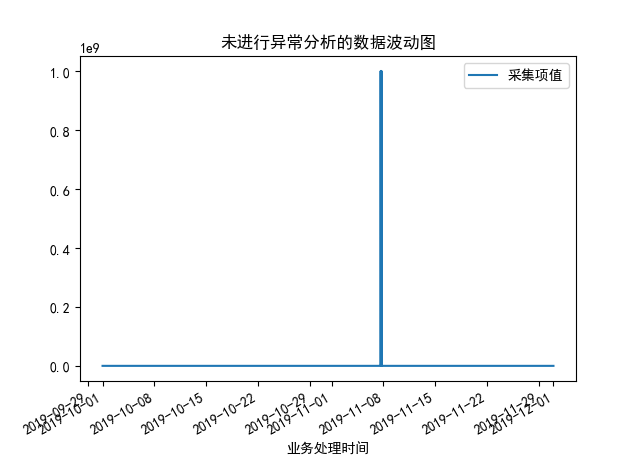
**三、初步方案**

**步骤1. 数据提取**

首先，从二级和三级数据库系统分别提取过程操作数据和相应的生产指标评价数据；其次，针对过程数据的多粒度（采样周期不同）特点，进行数据的对齐操作；然后，基于产品规格，使用大数据聚类算法，将数据分为不同的类别（或者工况）；最后，形成初始数据集合。

**步骤2. 数据清洗**

针对所形成的不同类别产品的初始数据集合，进行缺失数据的补全，以及异常数据的剔除，以保证数据的有效性，同时保证数据能够反映实际的生产规律变化情况。



**步骤3. 数据融合**

原始数据无统一的数据库平台，同平台下的数据表无明确分类，表中内容大多是数据信息的简单罗列，不能直接进行数据分析和建立模型，因此，需要对原数据进行数据融合，主要包括：

(1) 统一字段名称和类型；

(2) 对分散数据表按主题分类，对同一主题数据根据实际工艺操作再分类；

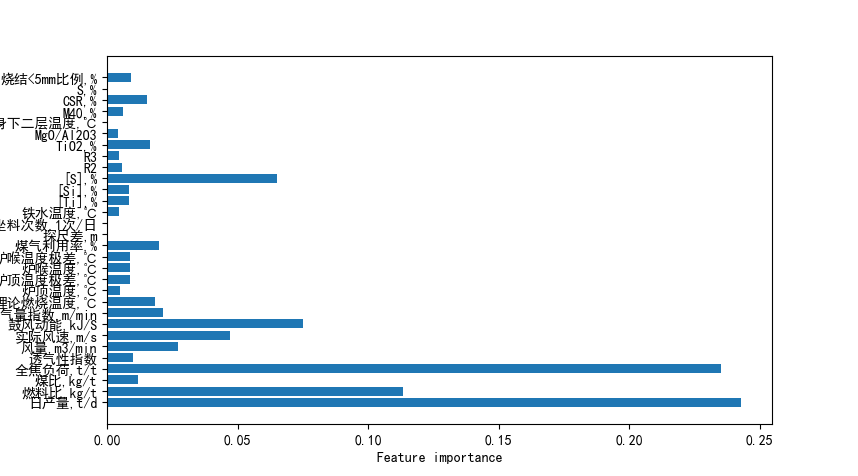
(3) 对字段进行重新组合，结合工艺设计新表，从而建立面向热镀锌操作经验提取的数据仓库。

**步骤4. 基于工艺经验筛选变量**

基于现场操作人员的工艺经验，初步确定与生产指标（产量、产品质量）相关的影响因素，这个范围可以稍大一些，并且简单给出这些影响因素之间的重要性排序。

**步骤5. 基于数据挖掘筛选变量**

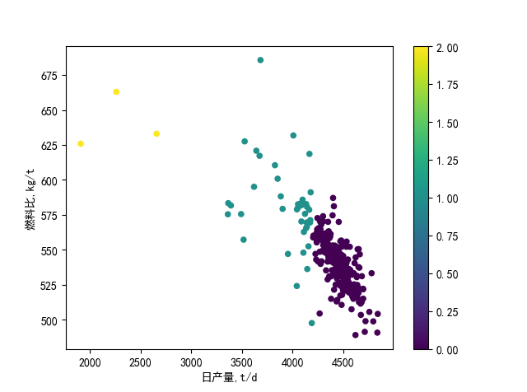
以可能影响生产指标（产量、产品质量）的相关工艺变量作为输入，以生产指标作为输出，基于大数据挖掘技术，对相关的工艺变量进行重要性排序，并且与人工经验进行的排序进行比较和讨论，最终确定出与生产指标最相关的工艺变量以及它们的重要性程度。（经过本步骤的数据挖掘，实际应用的关键工艺变量的数量相对于上一步骤中基于人工经验给出的变量数据，将会减少）



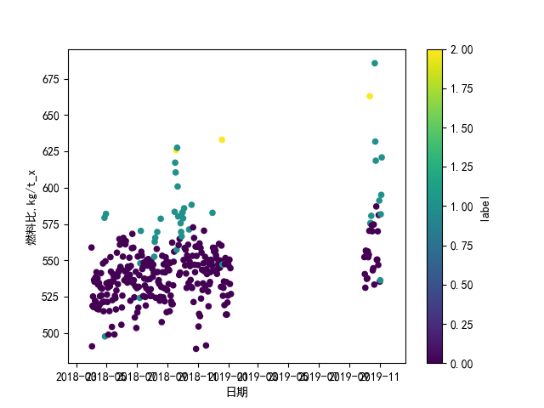
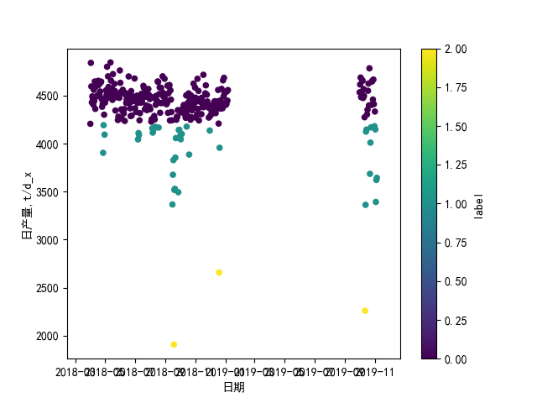
**步骤6. 操作经验大数据分析与多目标提取**

针对不同类别的带钢（工况），分别使用以下方法进行操作经验的大数据分析与多目标提取：

(1) 将标准化后的数据进行聚类分析，按照生产指标（产量、产品质量）分为不同的类别（推荐为3类：较好、较差、正常）



(2) 针对上一步骤所挖掘出的各关键工艺参数，利用聚类之后的数据进行多目标提取，以获得每个关键工艺参数的优化上限和下限。



(3) 基于以上的优化上下限，构建优化操作的星形图，以指导实际生产。

