题目：A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model

领域：银行内部数据 + 公共信誉记录（Public Credit Registers）来车贷行为评分模型

核心创新点：信誉的两个角度度量

论文结构与实现方法

* 数据处理：数据清洗、数据整合、数据转换、数据剔除和特征选择
  + 数据清洗：利用均值填补缺失值、平滑噪声数据、发现并剔除特异点，解决不一致性
  + 数据聚合：利用用户id连接不同数据集
  + 数据转换：分值数据正态化，根据逻辑继承关系数据数字化，数据通过分箱法
  + 数据剔除：分箱法
  + 特征选择：利用遗传算法，对特征进行排序（这里不是很详细，具体参见其中的引文）
* 分段分析：同质性分析，用户群体分类
* 一个角度度量的信誉度模型：针对每一个用户段，进行分析，调参，结果分十级
  + 逻辑回归+校准模型（calibration, 其实就是一个score的换算公式来对应一个G/b ODD RATIO 的东西，因为在文中对数据的处理用到了分段，G/b odd大概就是一个段占整体的比率）
* \*两个角度度量的信誉度模型：形成5\*5矩阵，利用K-S统计和ROC曲线分析结果
  + 评价标准：KS值（这个之后可用）、ROC曲线
  + 数据形态（其实挺重要的，但是全写上也没什么意义，就先不写了）
  + 其实所谓两个角度就是两个对应不同的数据集
* 信誉策略：对应矩阵，设计相应的信誉策略
  + 这个里面主要是针对情况对银行提建议，大概也就是分析人群的不同特点，然后用上面得到的分级去说明。

优秀之处：

数据集丰富，遗传算法的特征选择，数据预处理的分段

可采用的点：

数据预处理说全说细，也可以尝试分箱分段之类的，有一种先验知识的感觉。遗传算法的特征选择感觉不是很靠谱，但是以后可能特征选择可以进一步挖深。检验时使用了KS曲线可以尝试在之后使用