

因子切割方法论与遗传算法因子挖掘

演讲人：风哥

目录

- ① 因子切割方法论介绍与研究
- ② CTR因子切割实战
- ③ 遗传规划理论教学与实践

因子切割方法论介绍与研究

既然花式咖啡的成分里可以有苦有甜，那么涨跌幅的成分里为何不能区分出反转和动量？我们留意到，传统反转因子本质上是一段区间的涨跌幅，可以被很自然地拆分为许多更小的时段。那么，会不会存在这样的情况：组成传统反转因子的各个时段中，某些时段贡献了很强的反转，而某些时段只是贡献了很弱的反转、甚至是贡献了动量效应？换言之，信息在时间轴上的分布可能是不均匀的。

对象：具有可加性的目标变量

工具：有区分能力的切割指标

产出：对切割后变量的再加工

CTR因子切割实战

异质信念（Heterogeneous Belief）是由 Miller (1977) 1提出的一种存在于非完全理性的投资者间的行为，具体表现为卖空限制使悲观投资者被迫退出市场，从而导致股价系统地被高估。而 A 股是一个有着大量中小投资者且严格限制卖空的市场，市场上投资者并非都是理性投资者（对股票具有一致和准确的预期），同时由于存在信息不对称，因此不同投资者之间的预期是不同的（异质信念）。

异质信念的存在会导致股价被系统地高估，所对应的往往是个股层面的多空分歧，最直观的表现就是高换手率。低换手的股票是市场上未被系统高估的股票（处于低价位），能在未来获取较高的收益。因此我们认为换手率因子从某种程度上代表了该股票的“异质信念”。

遗传规划理论教学与实践

遗传算法是一种模仿自然选择和遗传机制的优化算法，主要通过模拟生物进化的过程来搜索解空间中的最优解。其核心原理包括遗传操作（选择、交叉和变异）和适应度评估。

1. 初始化种群：

在遗传算法开始时，首先需要随机生成一组初始解，称为种群。种群中的每个个体都代表了问题的一个可能解决方案。个体可以用染色体来表示，染色体由基因组成。这些解决方案可能是随机的，也可能基于先验知识生成。

2. 适应度评估：

对于种群中的每个个体，都需要计算其适应度值。适应度函数是衡量个体在解决问题中的优劣程度的函数。这个函数与问题的特性相关，目标是最大化或最小化适应度值。适应度值越高，表示个体的解决方案越好。

3. 选择：

选择是从当前种群中选择个体作为父代，用于产生下一代。选择的基本思想是根据个体的适应度值来确定被选中的概率，通常适应度更高的个体被选中的概率也更大。常见的选择方法包括轮盘赌选择、竞争选择和排名选择等。

4. 交叉：

在交叉操作中，选定的父代个体会以某种方式进行交叉，以产生子代个体。交叉操作模拟了生物学中的基因重组过程，通过交叉，子代个体可以获得父代个体的优良特征。常见的交叉方法包括单点交叉、多点交叉和均匀交叉等。

5. 变异:

变异操作是为了保持种群的多样性，防止算法陷入局部最优解。在变异阶段，个体的某些基因会以一定的概率发生变异，通常是随机的。变异操作使得搜索空间更广，有助于算法发现新的解。常见的变异操作包括单点变异、多点变异和均匀变异等。

6. 重复迭代:

以上步骤会循环迭代多次，直到满足停止条件。停止条件可以是达到最大迭代次数、找到满意解、适应度不再改善等。随着迭代的进行，种群中的个体逐渐演化，最终会收敛到一个或多个优秀的解。

欢迎提问
