Bayesian Optimization方法因子调优 总结报告

**一、贝叶斯优化原理**

At each step a Gaussian Process is fitted to the known samples (points previously explored), and the posterior distribution, combined with a exploration strategy (such as UCB (Upper Confidence Bound), or EI (Expected Improvement)), are used to determine the next point that should be explored.( <https://github.com/fmfn/BayesianOptimization>)

在使用此方法时，定义：

Target Function：可调参数为自变量，算出的IC为函数值；

Posterior Distribution：高斯分布；

Acquisition Function：UCB (Upper Confidence Bound)；

**二、在原有基础上的改进**

1. 优化过程参考fmfn/BayesianOptimization代码，但将其改为适用不同长度的因子（不同个数的可调参数）的程序，优化了接口；

2. 将具有固定模式的对应参数集中处理：

① 定义sum(X,N)/N中的两个N为同一个const\_node

② 定义regbeta(X,sequence(N),N)中的两个N为同一个const\_node

③ 限定sma(X,N,M)中的N>M

④ 限定std(X,N)中的N>=2

⑤ 限定corr(X,N)中的N>=2

3. 对INF和NaN值问题进行了优化，使其成为调优过程中的最坏情况；

4. 使用更好的IC计算方法：排序后用因子前40%和后40%的数据计算IC；

5. 防止过拟合情况出现作出的努力：

① 设置样本外对照组，作为因子调优过程中的重要依据

② 计算IC过程中删掉一些在时序上变化剧烈的点（未实行）

③ 计算IC过程中随机删除一些点（未实行）

**三、计算资源的占用情况**

1. 初始点设为10个；

2. 可能出现收敛情况的迭代次数：

|  |  |  |
| --- | --- | --- |
| 因子可调参数数 | 迭代次数 | 迭代时间 |
| 5个以下 | 70次左右 | 8min-15min |
| 5-7个 | 120次左右 | 20min-45min |
| 8-10个 | 180次左右 | 45min-70min |
| 10个以上 | 未探明 | 未探明 |

**四、因子回测分析**

1. 在Ann Return表现特别突出的因子：galpha191、galpha026、galpha153、galpha173、walpha044、walpha060 (Lt) galpha114 (Lb) 等，

其中galpha191：((corr(mean(volume,49),low,3)+((high+low)/2))-close)、galpha153：((((mean(close,99)+mean(close,12))+mean(close,59))+mean(close,48))/4)、walpha060：(0-(1\*((2\*scale(rank(((((close-low)-(high-close))/(high-low))\*volume))))-scale(rank(tsargmax(close,2))))))由于公式较短或者可调参数较少，在之前的调优方法中已经表现出色，贝叶斯优化方法体现不出优势；

而对于galpha026：(((sum(close,52)/52)-close)+corr(vwap,delay(close,8),29))、galpha173：((((3\*sma(close,92,8))-2)\*sma(sma(close,97,43),97,1))+sma(sma(sma(log(close),83,21),96,83),85,75))、galpha114：((rank(delay(((high-low)/(sum(close,72)/72)),1))\*rank(rank(volume)))/(((high-low)/(sum(close,1)/5))/(vwap-close)))这类公式相对复杂，且带有sum、sma导致参数比较多的情况时，往往表现优于之前的方法。

2. 表现取得较大提升的因子：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 因子名称 | M | IC | Sharpe | Ann Ret | MDD |
| tuned\_galpha026\_hs300\_bayes | Lt | 0.005 | 0.6394 | 0.1206 | 0.337 |
| tuned\_galpha114\_hs300\_bayes | Lb | -0.01 | 0.5011 | 0.1053 | 0.3272 |
| tuned\_walpha044\_hs300\_bayes | Lt | 0.0198 | 0.7687 | 0.13691 | 0.24271 |
| tuned\_galpha151\_hs300\_bayes | Lt | 0.0028 | 0.4139 | 0.0962 | 0.3477 |
| tuned\_galpha003\_hs300\_bayes | Lt | 0.0058 | 0.4235 | 0.0931 | 0.3384 |
| tuned\_galpha130\_hs300\_bayes | Lt | 0.0009 | 0.5041 | 0.0908 | 0.3294 |
| tuned\_galpha189\_hs300\_bayes | Lt | 0.0067 | 0.4246 | 0.0833 | 0.2467 |
| tuned\_galpha124\_hs300\_bayes | Lt | 0.0039 | 0.4751 | 0.0824 | 0.3474 |
| tuned\_galpha131\_hs300\_bayes | Lt | 0.0111 | 0.3095 | 0.067 | 0.2474 |
| tuned\_galpha129\_hs300\_bayes | Lt | 0.0081 | 0.2634 | 0.056 | 0.3517 |
| tuned\_galpha052\_hs300\_bayes | Lt | 0.0072 | 0.2592 | 0.053 | 0.2456 |
| tuned\_galpha052\_hs300\_bayes | Lt | 0.0072 | 0.2499 | 0.0468 | 0.2552 |
| tuned\_galpha054\_hs300\_bayes | Lt | 0.0117 | 0.2581 | 0.0455 | 0.2884 |
| tuned\_galpha059\_hs300\_bayes | Lt | 0.011 | 0.2001 | 0.0445 | 0.3622 |
| tuned\_galpha068\_hs300\_bayes | Lt | -0.003 | 0.2796 | 0.0428 | 0.331 |
| tuned\_galpha054\_hs300\_bayes | Lt | 0.0117 | 0.2124 | 0.0411 | 0.3038 |
| tuned\_galpha130\_hs300\_bayes | Lt | 0.0009 | 0.2443 | 0.0407 | 0.3195 |
| tuned\_galpha002\_hs300\_bayes | Lt | 0.0083 | 0.2004 | 0.0361 | 0.3428 |
| tuned\_galpha163\_hs300\_bayes | Lt | 0.0228 | 0.1654 | 0.0351 | 0.3183 |
| tuned\_galpha137\_hs300\_bayes | Lb | -0.0025 | 0.4954 | 0.0937 | 0.40713 |
| tuned\_galpha068\_hs300\_bayes | Lb | -0.0034 | 0.4165 | 0.0806 | 0.33985 |
| tuned\_galpha060\_hs300\_bayes | Lb | -0.0196 | 0.3776 | 0.0746 | 0.23998 |
| tuned\_galpha009\_hs300\_bayes | Lb | -0.0033 | 0.246 | 0.0508 | 0.32615 |
| tuned\_galpha074\_hs300\_bayes | Lb | -0.0108 | 0.2991 | 0.0503 | 0.30238 |
| tuned\_galpha114\_hs300\_bayes | Lb | -0.0099 | 0.2256 | 0.0448 | 0.32183 |
| tuned\_galpha098\_hs300\_bayes | Lb | -0.0072 | 0.1837 | 0.0373 | 0.30484 |
| tuned\_galpha077\_hs300\_bayes | Lb | -0.0046 | 0.222 | 0.035 | 0.30498 |
| tuned\_walpha050\_hs300\_bayes | Lt | 0.0118 | 0.2748 | 0.05161 | 0.33428 |
| tuned\_walpha015\_hs300\_bayes | Lt | 0.0151 | 0.2118 | 0.03636 | 0.34348 |
| tuned\_walpha052\_hs300\_bayes | Lb | -0.01 | 0.1813 | 0.0382 | 0.35195 |

其中参数较多的因子居多，也有一些短因子得到提升。

**五、得出的一些结论**

1. 此因子调优方法由于耗时较多，故应针对之前方法中表现不佳的因子进行优化；

2. 此方法可调出在某区间内的值，不必预设几个固定值；

3. 此方法对长因子、复杂公式较之前方法提升较大；

4. 此方法优化了对sum/corr/regbeta/sma等公式的适用。

**六、后续改进方向**

1. 对整体接口的优化，修复一些小Bug；

2. 贝叶斯优化过程中的一些参数还有改进的空间，Acquisition Function也不一定局限于UCB；

3. 过拟合问题仍需寻求更好的解决方案。