|  |
| --- |
| Table\_Title |
| **2018年07月05日** |
| **机器学习与量化投资：前沿研究之深度森林（gcForest）** |

|  |
| --- |
| Table\_BaseInfo |
| **金融工程主题报告** |
| 证券研究报告 |

|  |  |
| --- | --- |
| 杨勇 | 分析师 |
| SAC执业证书编号：S1450518010002  yangyong1@essence.com.cn | |
| 周袤 | 分析师 |
| SAC执业证书编号：S1450517120007  zhoumao@essence.com.cn | |

|  |
| --- |
| **Table\_Report** |
| **相关报告**   |  |  | | --- | --- | | 首只MSCI中国A股国际通指数增强基金发行，7月维持权益类风险判断 | 2018-07-02 | | 机器学习与CTA：数据挖掘与人类对世界的认识 | 2018-07-01 | | FOF和资产配置周报：创业板ETF持续流入，FOF净值破1 | 2018-06-24 | | 机器学习与CTA：商品期货策略继续表现出色 | 2018-06-24 | | 量化基金经理双周报：盈利和长期动量领先，市场重回核心资产 | 2018-06-20 | |

|  |
| --- |
| ■**gcForest算法：**gcForest（multi-Grained Cascade Forest）算法是2017年周志华教授提出来的一种基于树的深度模型，旨在作为深度神经网络的一种可供选择的替换。由于超参数更好的鲁棒性，小样本上更好的稳定性，因此该模型相对于神经网络可能在金融数据上有更好的表现。  ■**gcForest的回测表现：**将《机器学习与量化投资：避不开的那些事（1）》中的神经网络替换成为gcForest，按月收益回撤比可达15.959。  returns.png  ■**gcForest的参数敏感性：**该模型的各个参数的敏感性都非常低。  ■**风险提示：**  **根据历史信息及数据构建的模型在市场急剧变化时可能失效。** |

|  |
| --- |
| 内容目录 |
| [1. 综述 4](#_Toc518551334)  [2. 算法简介 4](#_Toc518551335)  [2.1. 决策树（decision tree） 4](#_Toc518551336)  [2.2. 随机森林（random forest） 4](#_Toc518551337)  [3. gcForest算法研究 4](#_Toc518551338)  [3.1. 多粒度扫描（multi grain scanning）阶段 4](#_Toc518551339)  [3.2. 级联森林（cascade forest structure）阶段 6](#_Toc518551340)  [3.3. 算法整体步骤 6](#_Toc518551341)  [4. gcForest算法在量化投资中的应用 7](#_Toc518551342)  [4.1. 方法 7](#_Toc518551343)  [4.1. 与随机森林的比较 8](#_Toc518551344)  [4.2. 参数敏感性分析 9](#_Toc518551345)  [4.2.1. n\_mgsRFtree=100与n\_mgsRFtree=150 10](#_Toc518551346)  [4.2.2. window=3与window=2 11](#_Toc518551347)  [4.2.3. n\_cascadeRF=2与n\_cascadeRF=1 12](#_Toc518551348)  [4.2.4. n\_cascadeRFtree=101与n\_cascadeRFtree=150 13](#_Toc518551349)  [4.2.5. min\_samples\_mgs=0.1与min\_samples\_mgs=0.15 15](#_Toc518551350)  [4.2.6. min\_samples\_cascade=0.1与min\_samples\_cascade=0.15 16](#_Toc518551351)  [4.2.7. 增幅临界值设定为2‰与增幅临界值设定为2.5‰ 17](#_Toc518551352) |
| 图表目录 |
| [图1：多粒度扫描（multi grain scanning）阶段 5](#_Toc518551274)  [图2：级联森林（cascade forest structure）阶段 6](#_Toc518551275)  [图3：gcForest算法框架 7](#_Toc518551276)  [图4：gcForest的参数介绍 7](#_Toc518551277)  [图5：gcForest的日净值曲线 8](#_Toc518551278)  [图6：随机森林的日净值曲线 8](#_Toc518551279)  [图7：gcForest的收益回撤比（月度） 9](#_Toc518551280)  [图8：随机森林的收益回撤比（月度） 9](#_Toc518551281)  [图9：gcForest的年度夏普比率 9](#_Toc518551282)  [图10：随机森林的年度夏普比率 9](#_Toc518551283)  [图11：n\_mgsRFtree=100的日净值曲线 10](#_Toc518551284)  [图12：n\_mgsRFtree=150的日净值曲线 10](#_Toc518551285)  [图13：n\_mgsRFtree=100的收益回撤比（月度） 10](#_Toc518551286)  [图14：n\_mgsRFtree=150的收益回撤比（月度） 10](#_Toc518551287)  [图15：n\_mgsRFtree=100的年度夏普比率 10](#_Toc518551288)  [图16：n\_mgsRFtree=150的年度夏普比率 10](#_Toc518551289)  [图17：window=3的日净值曲线 11](#_Toc518551290)  [图18：window=2的日净值曲线 11](#_Toc518551291)  [图19：window=3的收益回撤比（月度） 11](#_Toc518551292)  [图20：window=2的收益回撤比（月度） 11](#_Toc518551293)  [图21：window=3的年化夏普比率 12](#_Toc518551294)  [图22：window=2的年化夏普比率 12](#_Toc518551295)  [图23：4.2.3. n\_cascadeRF=2日净值曲线 12](#_Toc518551296)  [图24：n\_cascadeRF=1日净值曲线 12](#_Toc518551297)  [图25：n\_cascadeRF=2的收益回撤比（月度） 13](#_Toc518551298)  [图26：n\_cascadeRF=1的收益回撤比（月度） 13](#_Toc518551299)  [图27：n\_cascadeRF=2的年化夏普比率 13](#_Toc518551300)  [图28：n\_cascadeRF=1的年化夏普比率 13](#_Toc518551301)  [图29：n\_cascadeRFtree=101的日净值曲线 14](#_Toc518551302)  [图30：n\_cascadeRFtree=150的日净值曲线 14](#_Toc518551303)  [图31：n\_cascadeRFtree=101的收益回撤比（月度） 14](#_Toc518551304)  [图32：n\_cascadeRFtree=150的收益回撤比（月度） 14](#_Toc518551305)  [图33：n\_cascadeRFtree=101的年化夏普比率 14](#_Toc518551306)  [图34：n\_cascadeRFtree=150的年化夏普比率 14](#_Toc518551307)  [图35：min\_samples\_mgs=0.1的日净值曲线 15](#_Toc518551308)  [图36：min\_samples\_mgs=0.15的日净值曲线 15](#_Toc518551309)  [图37：min\_samples\_mgs=0.1的收益回撤比（月度） 15](#_Toc518551310)  [图38：min\_samples\_mgs=0.15的收益回撤比（月度） 15](#_Toc518551311)  [图39：min\_samples\_mgs=0.1的年化夏普比率 16](#_Toc518551312)  [图40：min\_samples\_mgs=0.15的年化夏普比率 16](#_Toc518551313)  [图41：min\_samples\_cascade=0.1的日净值曲线 16](#_Toc518551314)  [图42：min\_samples\_cascade=0.15的日净值曲线 16](#_Toc518551315)  [图43：min\_samples\_cascade=0.1的收益回撤比（月度） 17](#_Toc518551316)  [图44：min\_samples\_cascade=0.15的收益回撤比（月度） 17](#_Toc518551317)  [图45：min\_samples\_cascade=0.1的年化夏普比率 17](#_Toc518551318)  [图46：min\_samples\_cascade=0.15的年化夏普比率 17](#_Toc518551319)  [图47：增幅临界值设定为2‰的日净值曲线 18](#_Toc518551320)  [图48：增幅临界值设定为2.5‰的日净值曲线 18](#_Toc518551321)  [图49：增幅临界值设定为2‰的收益回撤比（月度） 18](#_Toc518551322)  [图50：增幅临界值设定为2.5‰的收益回撤比（月度） 18](#_Toc518551323)  [图51：增幅临界值设定为2‰的年化夏普比率 18](#_Toc518551324)  [图52：增幅临界值设定为2.5‰的年化夏普比率 18](#_Toc518551325) |
| [表1：gcForest与随机森林比较 9](#_Toc518551326)  [表2：\_mgsRFtree=100与n\_mgsRFtree=150比较 11](#_Toc518551327)  [表3：window=3与window=2比较 12](#_Toc518551328)  [表4：n\_cascadeRF=2与n\_cascadeRF=1比较 13](#_Toc518551329)  [表5：n\_cascadeRFtree=101与n\_cascadeRFtree=150比较 15](#_Toc518551330)  [表6：min\_samples\_mgs=0.1与min\_samples\_mgs=0.15比较 16](#_Toc518551331)  [表7：min\_samples\_cascade=0.1与min\_samples\_cascade=0.15比较 17](#_Toc518551332)  [表8：增幅临界值设定为2‰与增幅临界值设定为2.5‰比较 19](#_Toc518551333) |

1. 综述

gcForest（multi-Grained Cascade Forest）算法是2017年周志华教授提出来的一种基于树的深度模型，旨在作为深度神经网络的一种可供选择的替换。文章提出了该模型相比深度神经网络的诸多优势：不用大量地调参使得模型更容易训练，同样的超参数适用于各种不同的数据，在小规模数据也能有优秀的表现。该论文也进行了实验对比测试该算法的表现。本报告并非论证gcForest算法是否要比深度神经网络的表现更好，而是对此算法进行研究分析，介绍该算法的核心思想，以及它在量化投资决策中的表现性及参数敏感程度。

1. 算法简介
   1. 决策树（decision tree）

决策树是解决监督式机器学习的一类重要的方法，基于树结构的机器学习算法将特征空间划分成不交长方形（或长方体）的并，并且在每个小区域块中进行拟合，通常用一常数进行拟合。树的每个节点与某条垂直于特征空间坐标轴的（超）平面一一对应，此（超）平面所分割成的两部分分别由该节点的左右子决策树进行拟合。可以证明，在一棵决策树的某个叶节点上再次分割而产生的新决策树，它在训练集上的预测误差期望将不大于原决策树，即它在训练集上的学习效果将不劣于原决策树的学习效果。但叶节点包含的样本数过少，会降低小区域中拟合的可信度，进而产生过拟合现象。因此，叶节点包含样本数的最小值会作为决策树中的一个重要参数，控制决策树的学习率和过拟合程度。

* 1. 随机森林（random forest）

随机森林是由多棵决策树组成的结构，它将bagging的核心思想应用在了决策树的基本结构上，通过平均众多无偏有噪声的模型，得到无偏但方差更小的预测。容易证明，若， ， …， 独立同分布且，则; 若， ， …， 是同分布的且两两相关系数为ρ，且，则。为了真正减小平均后的方差，我们需要控制不同决策树之间的相关性。因此随机森林的做法是在训练集上进行bootstrapping，在每个bootstrapping产生的训练集上构建决策树，而树的每个节点先随机选择其中一部分特征（很多时候选择总特征数的平方根），并且在这些特征中选取一个最好的特征（可根据不同的准则）分割该节点。这样的做法大大增加了决策树的多样性，同时也有效的降低了不同决策树之间的相关性。值得注意的是，随机森林中树的棵树并非越大越好，因为树木数越多，不同决策树之间相关性增大的可能性越大。袋外误差（out-of-bag error）是随机森林的估计测试误差的一个重要指标，事实上，它是测试误差的无偏估计。

1. gcForest算法研究

gcForest是一种深度的森林模型，它主要分为多粒度扫描（multi grain scanning）阶段和级联森林（cascade forest structure）阶段。下面将详细介绍算法的原理。

* 1. 多粒度扫描（multi grain scanning）阶段

在训练过程中，它的多粒度扫描阶段将原特征用不同窗口进行切分得到大量新的特征，并且用这些新的特征训练不同的随机森林，并且将袋外决定函数（out of bag decision function）（即作为袋外数据计算所得概率分布）输入到级联森林训练阶段，用来训练该阶段的随机森林。在预测过程中，一个待预测的原始特征被同样切分后，输入到训练好的随机森林中进行预测，得到多个切片的分类概率分布，再作为整体输入到级联森林训练阶段，得到最终的预测。

|  |  |
| --- | --- |
| 图1：多粒度扫描（multi grain scanning）阶段 |  |
| Screen Shot 2018-04-27 at 9.58.36 AM.png |  |
| 资料来源：Deep Forest: Towards An Alternative to Deep Neural Networks，安信证券研究中心 |  |

如上图，一个 400 维的序列特征通过不断向下滑动 100 维窗口，可以获得 301 个 100 维的特征； 一个 20\*20 的图像特征被10\*10的窗口不断滑动，可以获得 121 个 10\*10 的图像。将这些已经切分好的特征的输出变量仍标记为原特征的输出变量，作为新的训练集。用来训练随机森林。因此对400维的序列，则训练集将被扩大301倍，对20\*20的图像，则训练集将被扩大121倍。图 1 所示例子为三分类问题，两个独立随机森林的情形，则袋外决定函数矩阵在的每个训练样本上均得到⼀个维数为3的概率分布，再按训练样本将切分后的训练样本得到的袋外决定函数合并起来得到下⼀阶段的训练集，如上图，将每个 400维原始特征变成 1806 维特征；或将每个 20\*20 的特征变成 726维特征。

多粒度扫描阶段最大的贡献是产生了更多的特征，它的主要思想来自多样例学习（multiple instance learning)。一个生动的例子是：我们想要让机器学习一张图是否为“海滩”，我们可以将任务分解为这张图是否包含“沙子”和“水”。抽象而言，学习者接收到一系列被标记的包，每个包中包含了一部分样例，如果一个包中的样例均被标记为“负”，则这个包被标记为“负”；如果一个包中的样例至少包含一个样例被标记为“正”，则这个包被标记为“正”。在多粒度扫描阶段，每个切分过的窗口均被标记为它应属的归类，用来训练随机森林。试想这些切片中有有一片的图像为“水”，被标记为“海滩”，有一片的图像为“沙子”，被标记为“海滩”，均用来训练多粒度扫描阶段的随机森林。袋外决定函数会给出这样的结果：如果这一片图像为“水”，那么它有一定概率为“海滩”，因为它学习的其他图像中也可能包含水但未被标记为“海滩”；同样如果这一片图像为“沙子”，则它也有一定概率为“海滩”；如果某一片图像中为“汽车”，则它会有很小概率为“海滩”，因为“汽车”并非“海滩”的必要条件，且“海滩”的训练集中几乎不会出现“汽车”。因此每一个切片都被赋予了是否为“海滩”的概率输入到了级联森林训练中，让这一阶段的随机森林判断这些切片结合起来是否为“海滩”。因此多粒度扫描阶段可以更好地从大量新的特征中识别出有效特征，来进行下一步的预测。

* 1. 级联森林（cascade forest structure）阶段

训练过程的第二个阶段是级联森林训练阶段，首先将训练集按比例分成两部分：训练集和验证集。用训练集训练第一层的随机森林，此时将验证集输入到这一层的随机森林中测试准确率，然后将袋外决定函数与原来的训练集结合起来作为下一层的训练集，训练下一层的随机森林，并且继续测试其在验证集上的准确率，若准确率提升低于容忍度（gcForest参数之一），则层级停止生长，否则将会一直继续下去，直到层级大于某一预先设定值（gcForest参数之一）。在预测过程中，输入特征向量依次进入每一层的随机森林，将得到的预测概率分布与原输入特征结合起来进入下一层的随机森林，最后将最后一层的随机森林得到的预测概率分布平均，取最大值对应的分类为最终的预测分类。

|  |
| --- |
| 图2：级联森林（cascade forest structure）阶段 |
| Screen Shot 2018-04-27 at 9.58.10 AM.png |
| 资料来源：Deep Forest: Towards An Alternative to Deep Neural Networks，安信证券研究中心 |

这一阶段的思想在于模仿深度神经网络中深度的概念，又使其自适应的生长层级。虽然gcForest模型中“深度”带来的收益可能不如深度神经网络，但是此模型打破了深度模型只适用于神经网络的局限。

* 1. 算法整体步骤

如下面算法框架所示，第1-14行为训练阶段，用来训练样本（X, y）；第15-20行为预测阶段，给出待预测集Xpredict的预测分类ypredict。具体而言，算法第1-3行为训练阶段的多粒度扫描阶段，Xmgs为多粒度扫描阶段的输出结果，在图1中为序列数据产生的1806维或图像数据产生的的726维向量；算法第4-14行为训练阶段的级联森林阶段，得到共有maxlayer层已训练好的随机森林；第15-16行为预测阶段的多粒度扫描阶段，Xmgs’为输出结果，同样如图1中所示；第17-20行为训练阶段的级联森林阶段，如图2，Xmgs’为input feature vector，输入到训练好的级联森林，得到最终预测final prediction。

|  |
| --- |
| 图3：gcForest算法框架 |
| algorithm.png |
| 资料来源：安信证券研究中心整理 |

1. gcForest算法在量化投资中的应用
   1. 方法

除了用gcForest的方法替换神经网络以外，其他方法都与《机器学习与量化投资：避不开的那些事（1）》一致。

以中证500指数为例，我们在10：00，10：30，11：00，13：01，13：30，14：00，14：30这7个时间点对半小时后的close做出预测，并根据此预测做出仓位决策。若预测的close的增幅大于2‰，则设置仓位为1；若预测的close的增幅小于-2‰，则设置仓位为-1；否则保持原来仓位。我们将gcForest算法的参数设置为如下：shape\_1X对不同特征取值不同，n\_mgsRFtree=100，window=3，stride=1，cascade\_test\_size=0.2，n\_cascadeRF=2，n\_cascadeRFtree=101，min\_samples\_mgs=0.1，min\_samples\_cascade=0.1，cascade\_layer=np.inf，tolerance=0.0，n\_jobs=1。

|  |
| --- |
| 图4：gcForest的参数介绍 |
| Screen Shot 2018-05-02 at 12.43.48 PM.png |
| 资料来源：安信证券研究中心整理 |

* 1. 与随机森林的比较

首先我们比较gcForest与随机森林的表现。

|  |  |  |
| --- | --- | --- |
| 图5：gcForest的日净值曲线 |  | 图6：随机森林的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图7：gcForest的收益回撤比（月度） |  | 图8：随机森林的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图9：gcForest的年度夏普比率 |  | 图10：随机森林的年度夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表1：gcForest与随机森林比较

|  |  |  |
| --- | --- | --- |
|  | gcForest | 随机森林 |
| **收益回撤比（按月统计）** | 15.959 | 6.0397 |
| **盈亏比** | 1.326 | 1.163 |
| **平均持仓时间** | 149.441 | 113.692 |
| **单笔平均收益** | 0.00165 | 0.000616 |
| **单笔最大收益** | 0.110 | 0.150 |
| **单笔最大损失** | -0.120 | -0.123 |

资料来源：安信证券研究中心整理

* 1. 参数敏感性分析

参数敏感性分析有利于我们建立对模型的信心。一个非常参数敏感的模型，有可能是模型过度拟合的一个标志，在实盘时候会有很大的风险。由于gcForest是一个全新的算法，仔细的研究参数敏感性将非常重要。

通过以下分析可以看出，除了划分1，-1, 0的标准（也即是大于2‰贴标签为1，小于-2‰贴标签为-1，不然则为0）会对模型产生比较严重的影响之外，其他参数对模型的影响微乎其微。

* + 1. n\_mgsRFtree=100与n\_mgsRFtree=150

|  |  |  |
| --- | --- | --- |
| 图11：n\_mgsRFtree=100的日净值曲线 |  | 图12：n\_mgsRFtree=150的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图13：n\_mgsRFtree=100的收益回撤比（月度） |  | 图14：n\_mgsRFtree=150的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图15：n\_mgsRFtree=100的年度夏普比率 |  | 图16：n\_mgsRFtree=150的年度夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表2：\_mgsRFtree=100与n\_mgsRFtree=150比较

|  |  |  |
| --- | --- | --- |
|  | n\_mgsRFtree=100 | n\_mgsRFtree=150 |
| **收益回撤比（按月统计）** | 15.959 | 14.386 |
| **盈亏比** | 1.326 | 1.219 |
| **平均持仓时间** | 149.441 | 151.361 |
| **单笔平均收益** | 0.00165 | 0.00151 |
| **单笔最大收益** | 0.110 | 0.126 |
| **单笔最大损失** | -0.120 | -0.120 |

资料来源：安信证券研究中心整理

* + 1. window=3与window=2

|  |  |  |
| --- | --- | --- |
| 图17：window=3的日净值曲线 |  | 图18：window=2的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图19：window=3的收益回撤比（月度） |  | 图20：window=2的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图21：window=3的年化夏普比率 |  | 图22：window=2的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表3：window=3与window=2比较

|  |  |  |
| --- | --- | --- |
|  | window=3 | window=2 |
| **收益回撤比（按月统计）** | 15.959 | 19.723 |
| **盈亏比** | 1.326 | 1.294 |
| **平均持仓时间** | 149.441 | 159.792 |
| **单笔平均收益** | 0.00165 | 0.00169 |
| **单笔最大收益** | 0.110 | 0.106 |
| **单笔最大损失** | -0.120 | -0.120 |

资料来源：安信证券研究中心整理

* + 1. n\_cascadeRF=2与n\_cascadeRF=1

|  |  |  |
| --- | --- | --- |
| 图23：4.2.3. n\_cascadeRF=2日净值曲线 |  | 图24：n\_cascadeRF=1日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图25：n\_cascadeRF=2的收益回撤比（月度） |  | 图26：n\_cascadeRF=1的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图27：n\_cascadeRF=2的年化夏普比率 |  | 图28：n\_cascadeRF=1的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表4：n\_cascadeRF=2与n\_cascadeRF=1比较

|  |  |  |
| --- | --- | --- |
|  | n\_cascadeRF=2 | n\_cascadeRF=1 |
| **收益回撤比** | 15.959 | 26.651 |
| **盈亏比** | 1.326 | 1.293 |
| **平均持仓时间** | 149.441 | 151.279 |
| **单笔平均收益** | 0.00165 | 0.00189 |
| **单笔最大收益** | 0.110 | 0.145 |
| **单笔最大损失** | -0.120 | -0.0985 |

资料来源：安信证券研究中心整理

* + 1. n\_cascadeRFtree=101与n\_cascadeRFtree=150

|  |  |  |
| --- | --- | --- |
| 图29：n\_cascadeRFtree=101的日净值曲线 |  | 图30：n\_cascadeRFtree=150的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图31：n\_cascadeRFtree=101的收益回撤比（月度） |  | 图32：n\_cascadeRFtree=150的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图33：n\_cascadeRFtree=101的年化夏普比率 |  | 图34：n\_cascadeRFtree=150的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表5：n\_cascadeRFtree=101与n\_cascadeRFtree=150比较

|  |  |  |
| --- | --- | --- |
|  | n\_cascadeRFtree=101 | n\_cascadeRFtree=160 |
| **收益回撤比** | 15.959 | 19.131 |
| **盈亏比** | 1.326 | 1.249 |
| **平均持仓时间** | 149.441 | 150.825 |
| **单笔平均收益** | 0.00165 | 0.00157 |
| **单笔最大收益** | 0.110 | 0.152 |
| **单笔最大损失** | -0.120 | -0.120 |

资料来源：安信证券研究中心整理

* + 1. min\_samples\_mgs=0.1与min\_samples\_mgs=0.15

|  |  |  |
| --- | --- | --- |
| 图35：min\_samples\_mgs=0.1的日净值曲线 |  | 图36：min\_samples\_mgs=0.15的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图37：min\_samples\_mgs=0.1的收益回撤比（月度） |  | 图38：min\_samples\_mgs=0.15的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图39：min\_samples\_mgs=0.1的年化夏普比率 |  | 图40：min\_samples\_mgs=0.15的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表6：min\_samples\_mgs=0.1与min\_samples\_mgs=0.15比较

|  |  |  |
| --- | --- | --- |
|  | min\_samples\_mgs=0.1 | min\_samples\_mgs=0.15 |
| **收益回撤比** | 15.959 | 20.799 |
| **盈亏比** | 1.326 | 1.327 |
| **平均持仓时间** | 149.441 | 148.959 |
| **单笔平均收益** | 0.00165 | 0.00175 |
| **单笔最大收益** | 0.110 | 0.112 |
| **单笔最大损失** | -0.120 | -0.100 |

资料来源：安信证券研究中心整理

* + 1. min\_samples\_cascade=0.1与min\_samples\_cascade=0.15

|  |  |  |
| --- | --- | --- |
| 图41：min\_samples\_cascade=0.1的日净值曲线 |  | 图42：min\_samples\_cascade=0.15的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图43：min\_samples\_cascade=0.1的收益回撤比（月度） |  | 图44：min\_samples\_cascade=0.15的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图45：min\_samples\_cascade=0.1的年化夏普比率 |  | 图46：min\_samples\_cascade=0.15的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表7：min\_samples\_cascade=0.1与min\_samples\_cascade=0.15比较

|  |  |  |
| --- | --- | --- |
|  | min\_samples\_cascade=0.1 | min\_samples\_cascade=0.15 |
| **收益回撤比** | 15.959 | 18.741 |
| **盈亏比** | 1.326 | 1.299 |
| **平均持仓时间** | 149.441 | 162.703 |
| **单笔平均收益** | 0.00165 | 0.00183 |
| **单笔最大收益** | 0.110 | 0.155 |
| **单笔最大损失** | -0.120 | -0.123 |

资料来源：安信证券研究中心整理

* + 1. 增幅临界值设定为2‰与增幅临界值设定为2.5‰

|  |  |  |
| --- | --- | --- |
| 图47：增幅临界值设定为2‰的日净值曲线 |  | 图48：增幅临界值设定为2.5‰的日净值曲线 |
| returns.png |  | returns.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图49：增幅临界值设定为2‰的收益回撤比（月度） |  | 图50：增幅临界值设定为2.5‰的收益回撤比（月度） |
| calmar ratio.png |  | calmar ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

|  |  |  |
| --- | --- | --- |
| 图51：增幅临界值设定为2‰的年化夏普比率 |  | 图52：增幅临界值设定为2.5‰的年化夏普比率 |
| sharpe ratio.png |  | sharpe ratio.png |
| 资料来源：安信证券研究中心整理 |  | 资料来源：安信证券研究中心整理 |

表8：增幅临界值设定为2‰与增幅临界值设定为2.5‰比较

|  |  |  |
| --- | --- | --- |
|  | 增幅临界值设定为2‰ | 增幅临界值设定为2.5‰ |
| **收益回撤比** | 15.959 | 15.149 |
| **盈亏比** | 1.326 | 1.216 |
| **平均持仓时间** | 149.441 | 221.561 |
| **单笔平均收益** | 0.00165 | 0.00177 |
| **单笔最大收益** | 0.110 | 0.181 |
| **单笔最大损失** | -0.120 | -0.196 |

资料来源：安信证券研究中心整理

实习生张明睿对此文有贡献

|  |
| --- |
| **Ta**ble\_AuthorStatement |
| * **分析师声明** |
| |  | | --- | | 杨勇、周袤声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。 | |

|  |
| --- |
| * **本公司具备证券投资咨询业务资格的说明** |
| 安信证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。 |

|  |
| --- |
| * **免责声明** |
| 本报告仅供安信证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。  本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准，如有需要，客户可以向本公司投资顾问进一步咨询。  在法律许可的情况下，本公司及所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。  本报告版权仅为本公司所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发或引用本报告的任何部分。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“安信证券股份有限公司研究中心”，且不得对本报告进行任何有悖原意的引用、删节和修改。  本报告的估值结果和分析结论是基于所预定的假设，并采用适当的估值方法和模型得出的，由于假设、估值方法和模型均存在一定的局限性，估值结果和分析结论也存在局限性，请谨慎使用。  安信证券股份有限公司对本声明条款具有惟一修改权和最终解释权。 |

|  |
| --- |
| **Table\_Sales** |
| * **销售联系人**  |  |  |  |  | | --- | --- | --- | --- | | 上海联系人 | 朱贤 | 021-35082852 | zhuxian@essence.com.cn | |  | 孟硕丰 | 021-35082788 | mengsf@essence.com.cn | |  | 李栋 | 021-35082821 | lidong1@essence.com.cn | |  | 侯海霞 | 021-35082870 | houhx@essence.com.cn | |  | 林立 | 021-68766209 | linli1@essence.com.cn | |  | 潘艳 | 021-35082957 | panyan@essence.com.cn | |  | 刘恭懿 | 021-35082961 | liugy@essence.com.cn | |  | 孟昊琳 | 021-35082963 | menghl@essence.com.cn | | 北京联系人 | 温鹏 | 010-83321350 | wenpeng@essence.com.cn | |  | 田星汉 | 010-83321362 | tianxh@essence.com.cn | |  | 王秋实 | 010-83321351 | wangqs@essence.com.cn | |  | 张莹 | 010-83321366 | zhangying1@essence.com.cn | |  | 李倩 | 010-83321355 | liqian1@essence.com.cn | |  | 姜雪 | 010-59113596 | jiangxue1@essence.com.cn | |  | 王帅 | 010-83321351 | wangshuai1@essence.com.cn | | 深圳联系人 | 胡珍 | 0755-82558073 | huzhen@essence.com.cn | |  | 范洪群 | 0755-82558044 | fanhq@essence.com.cn | |  | 杨晔 | 0755-82558046 | yangye@essence.com.cn | |  | 巢莫雯 | 0755-82558183 | chaomw@essence.com.cn | |  | 王红彦 | 0755-82558361 | wanghy8@essence.com.cn | |  | 黎欢 | 0755-82558045 | lihuan@essence.com.cn | |

|  |  |
| --- | --- |
| Table\_Address | |
| **安信证券研究中心** | |
| **深圳市** | |
| **地 址：** | **深圳市福田区深南大道2008号中国凤凰大厦1栋7层** |
| **邮 编：** | 518026 |
| **上海市** | |
| **地 址：** | **上海市虹口区东大名路638号国投大厦3层** |
| **邮 编：** | 200080 |
| **北京市** | |
| **地 址：** | **北京市西城区阜成门北大街2号楼国投金融大厦15层** |
| **邮 编：** | 100034 |