50支股票上的聚类，相关度的平均值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Window | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 |
| HTM1 | 0.5548 | 0.6126 | 0.6287 | 0.6330 | 0.6342 | 0.6348 | 0.6345 | 0.6343 |
| Expert1 | 0.5172 | 0.5747 | 0.5879 | 0.5915 | 0.5927 | 0.5934 | 0.5930 | 0.5925 |
| HTM2 | 0.5266 | 0.5829 | 0.6133 | 0.6372 | 0.6554 | 0.6545 | 0.6542 | 0.6552 |
| Expert2 | 0.4948 | 0.5605 | 0.5927 | 0.6154 | 0.6314 | 0.6303 | 0.6285 | 0.6284 |

全部2209支股票上的聚类，相关度的平均值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| window | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 |
| HTM1 | 0.4944 | 0.5265 | 0.5354 | 0.5377 | 0.5396 | 0.5403 | 0.5402 | 0.5402 |
| Rnd1 | 0.3599 | 0.3934 | 0.4027 | 0.4053 | 0.4069 | 0.4064 | 0.4051 | 0.4040 |
| HTM2 | 0.4446 | 0.4941 | 0.5175 | 0.5354 | 0.5459 | 0.5438 | 0.5415 | 0.5407 |
| Rnd2 | 0.3436 | 0.4035 | 0.4301 | 0.4490 | 0.4625 | 0.4593 | 0.4564 | 0.4559 |

相关度计算方法：

以3、5、……、17天为窗口大小，计算一个类中所有股票对的价格之间的皮尔森系数，将它们的平均值取绝对值后作为该类的相关度值。对所有聚类产生的类都作如上操作，得到的相关度值的平均值。

HTM1、HTM2表示评价的是我们的方法在第一段（2012/5/4-2012/9/27）、第二段（2012/9/27-2012/12/31）时间上的聚类结果；Expert1、2为相应的根据板块计算出来的结果。Rnd1、Rnd2是随机地产生聚类，10次的平均值。

去除单个元素的聚类的方法：计算所有非单个元素的聚类各自在price data上的平均值（中心），然后计算单元素类中那个股票的price data与各类中心的距离，取最近的一个。

随着聚类数量的减少，会使得上面的相关度变小，当然purity也会变小，nmi的变化不一定。