The authors repeat each experiment 30 times, but they do not perform any statistical test to assess the significance of the differences between RETECS and PA. In RQ1, the differences between the two approaches are very slight and the claim that PA outperforms RETECS is not adequately supported without hypothesis testing.

思路：使用t分布检验RETECS方法和PA方法在NAPFD、RECALL、TTF指标上是否具有显著性差异。

其中显著性水平a=0.01,n=30。

H0:RETECS方法和PA方法在NAPFD、RECALL、TTF指标上不具有显著性差异。

H1:RETECS方法和PA方法在NAPFD、RECALL、TTF指标上具有显著性差异。

具体方法：使用scipy.stas.ttest\_ind函数对30次实验的NAPFD值计算p值。例如dspace数据集上计算NAPFD的显著性差异，RETECS方法重复30次得到30个NAPFD值，PA方法重复30次得到30个NAPFD值，传入scipy.stas.ttest\_ind即可求出P值。

RETECS和PA方法在NAPFD,RECALL,TTF三个指标上的p值表，a=0.01,灰色表示p>a

|  |  |  |  |
| --- | --- | --- | --- |
|  | napfd | recall | ttf |
| apache\_commons | 2.65E-01 | 1.00E+00 | 6.67E-01 |
| apache\_drill | 9.72E-01 | 6.99E-01 | 8.23E-01 |
| apache\_parquet | 5.27E-03 | 2.26E-01 | 3.61E-01 |
| apache\_tajo | 9.88E-02 | 7.21E-10 | 8.59E-08 |
| dspace | 9.46E-07 | 3.02E-07 | 1.56E-04 |
| google\_auto | 1.82E-01 | 8.76E-01 | 2.91E-02 |
| google\_closure | 9.63E-03 | 2.63E-01 | 2.75E-01 |
| google\_guava | 2.09E-01 | 1.82E-01 | 1.99E-01 |
| iofrol | 2.93E-01 | 2.00E-01 | 6.44E-01 |
| mybatis | 5.76E-02 | 9.16E-02 | 2.05E-01 |
| paintcontrol | 4.19E-01 | 3.67E-01 | 7.92E-01 |
| rails | 5.13E-04 | 2.46E-01 | 8.49E-02 |

通过上表，在NAPFD上，我们假设犯错概率为a=0.01，貌似要接受H0了，也就是显著性差异确实不大。

a=0.3左右我们才能拒绝H0,