候选值 10个

各20组实验

\*取真值概率为采样概率

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 取真值概率 | MajorityVoting | TruthFinder | 2ESTIMATES | 3ESTIMATES | Investment |
| 10% | 9.4 | 9.9 | 9.5 | 8.4 | 10.05 |
| 20% | 31.85 | 33.35 | 32.95 | 30.4 | 33.45 |
| 30% | 57.95 | 59.9 | 59.2 | 54.4 | 59.6 |
| 40% | 81.4 | 82.2 | 82.3 | 77.3 | 81.8 |
| 50% | 93.25 | 93.0 | 93.45 | 89.3 | 92.2 |
| 60% | 97.85 | 97.6 | 97.5 | 95.1 | 97.2 |
|  |  |  |  |  |  |

3ESTIMATES果然是我实现的不对……在所有实验中都表现出了最差的性能。

我们很高兴的看到在取真值的概率没有那么大（数据没有那么好）的情况下，大多数算法都表现出了优于MajorityVoting的性能，分界线出现在50%，在50%取真值概率的情况下，除了2ESTIMATES之外其他的所有算法都不如majorityvoting的性能。在取真值概率大于50%时，MajorityVoting的优势进一步被放大，而小于该值时，算法都比MajorityVoting要好。

分析原因是因为在50%取真值的情况下，由于候选值有足足10个，导致几乎大部分情况下都会给出正确答案，也即数据质量太好了。在这种情况下，我们可以认为每一个source都是优质的，那么他们做出的投票结果也将是优质的。