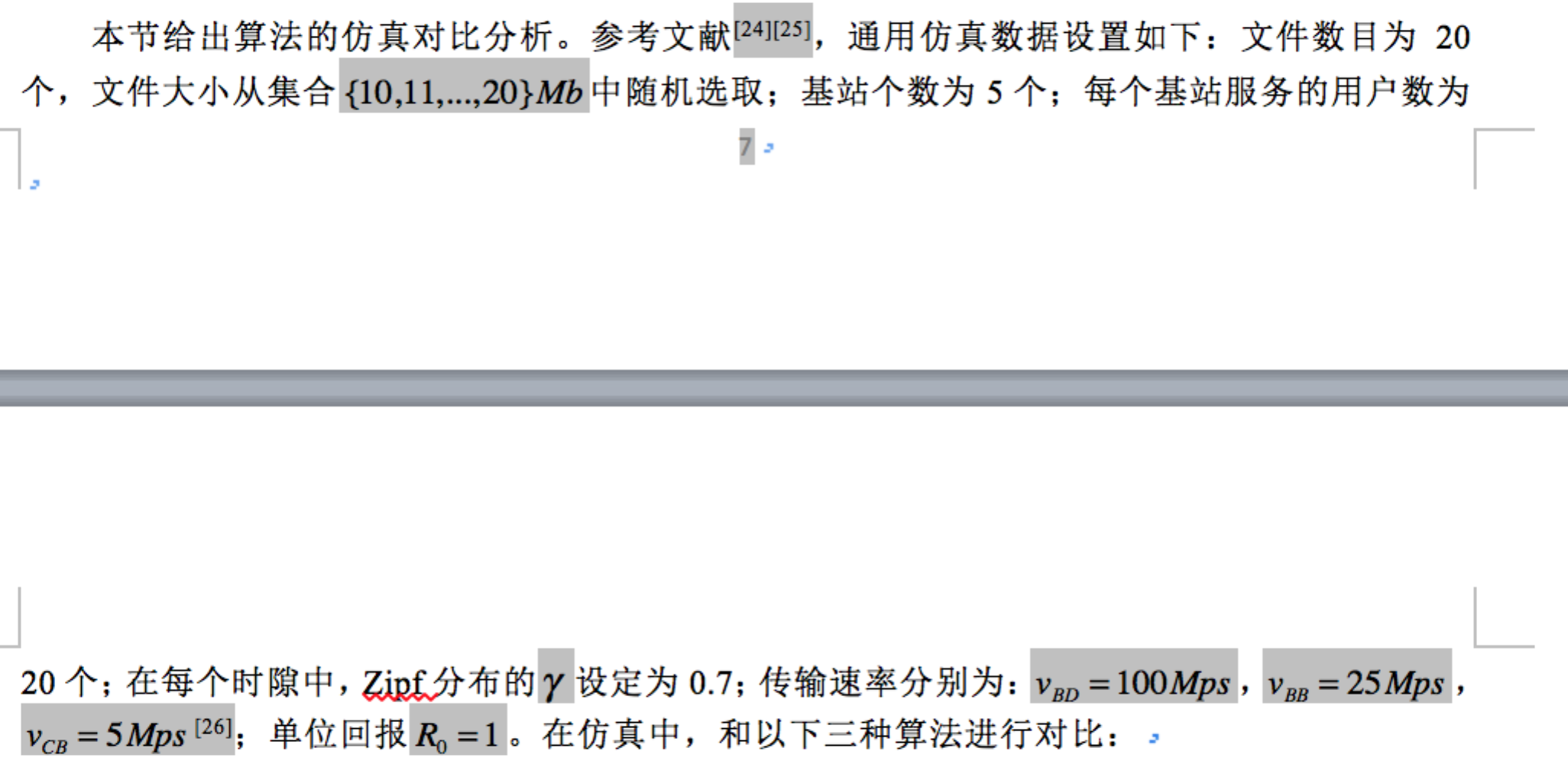
|  |  |
| --- | --- |
| 基站(B) | 5 |
| 文件(K) | 20 |
| 存储(M) | 100 |
| Zipf\_(delta) | 两状态markov chain  状态：states = [0.9, 1.2]  转移矩阵：transition=[[0.75, 0.25],  [0.25, 0.75]] |
| \Epsilon\_t | 0.05 |
| 惩罚项系数 \beta | 1 |
| 学习率 \alpha\_t | 0.1 |



1. markov\_state.jpg: 两状态 markov 流行度状态累计平均概率图：t 时刻的值代表从1时刻到 t 时刻该状态出现的平均概率图。特意选取状态转移对称的转移概率，刻画流行度的变化更为合理。
2. algorithms\_comparison.jpg: 实时回报R\_t 对比图。
3. q\_values.jpg: 算法 Q-value 迭代图，不同学习率 Q-value 迭代对比图。
4. caching\_rate.jpg: 累计平均缓存命中率图：t 时刻的值代表从 1时刻到 t 时刻算法的平均缓存命中率。
5. network\_topology.jpg: 大规模网络拓扑图，红点代表宏基站，圆圈代表宏基站的服务范围。
6. distributed\_q\_learning.jpg: 分布式 Q-learning 算法图。双轴，左轴线代表20个基站的实时 reward 迭代图，对应20无 marker 线；右轴线代表实时的reward 方差图，对应黑色 marker 线，方差的迭代减小表明了算法的聚合性。