

6. Results

| Task Id | Name | Hop 2 | Hop 3 |
|---------|----------------------|-------|-------|
| 1 | 1 supporting fact | 1.0 | 0.99 |
| 2 | 2 supporting facts | 0.86 | 0.86 |
| 3 | 3 supporting facts | 0.38 | 0.36 |
| 4 | 2 argument relations | 0.98 | 0.973 |
| 5 | 3 argument relations | 0.86 | 0.88 |
| 6 | yes/no questions | 0.94 | 0.93 |
| 7 | counting | 0.85 | 0.88 |
| 8 | lists/sets | 0.88 | 0.88 |
| 9 | simple negation | 0.7 | 0.73 |
| 10 | indefinite knowledge | 0.79 | 0.78 |
| 11 | basic coreference | 0.83 | 0.84 |
| 12 | conjunction | 0.99 | 1.0 |
| 13 | compound coreference | 0.86 | 0.88 |
| 14 | time reasoning | 0.9 | 0.89 |
| 15 | basic deduction | 0.92 | 1.0 |
| 16 | basic induction | 0.44 | 0.442 |
| 17 | positional reasoning | 0.51 | 0.5 |
| 18 | size reasoning | 0.836 | 0.85 |
| 19 | path finding | 0.134 | 0.12 |
| 20 | agent's motivation | 1.0 | 1.0 |
| | Accuracy > 0.95 | 5 | 6 |

Accuracy by tasks

