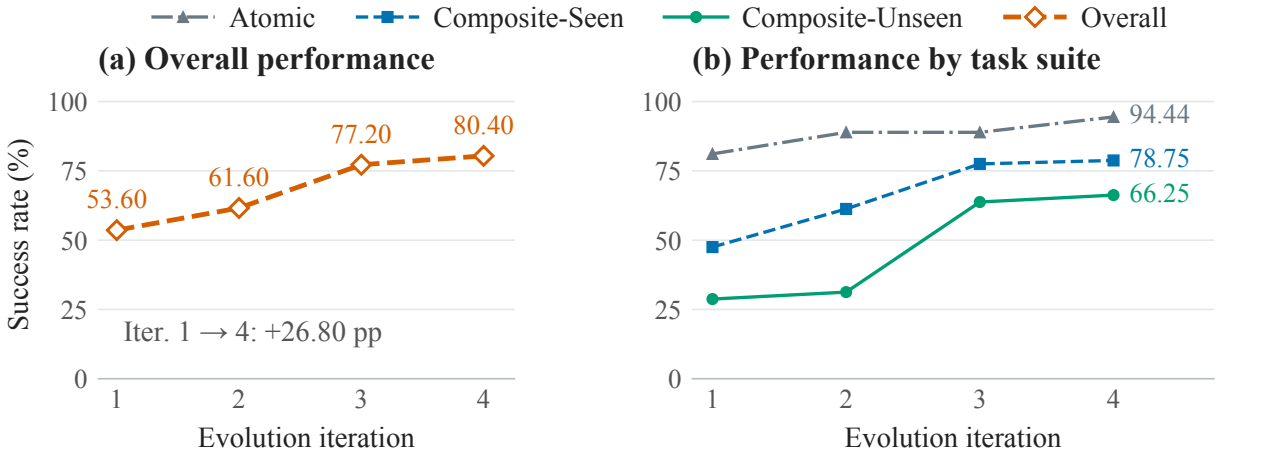
Atomic
Composite-Seen
Composite-Unseen
Overall
(a) Overall performance
Success rate (%)
100
75
50
25
0
53.60
61.60
77.20
80.40
Iter. 1 → 4: +26.80 pp
1
2
3
4
Evolution iteration
(b) Performance by task suite
94.44
78.75
66.25
Evolution iteration