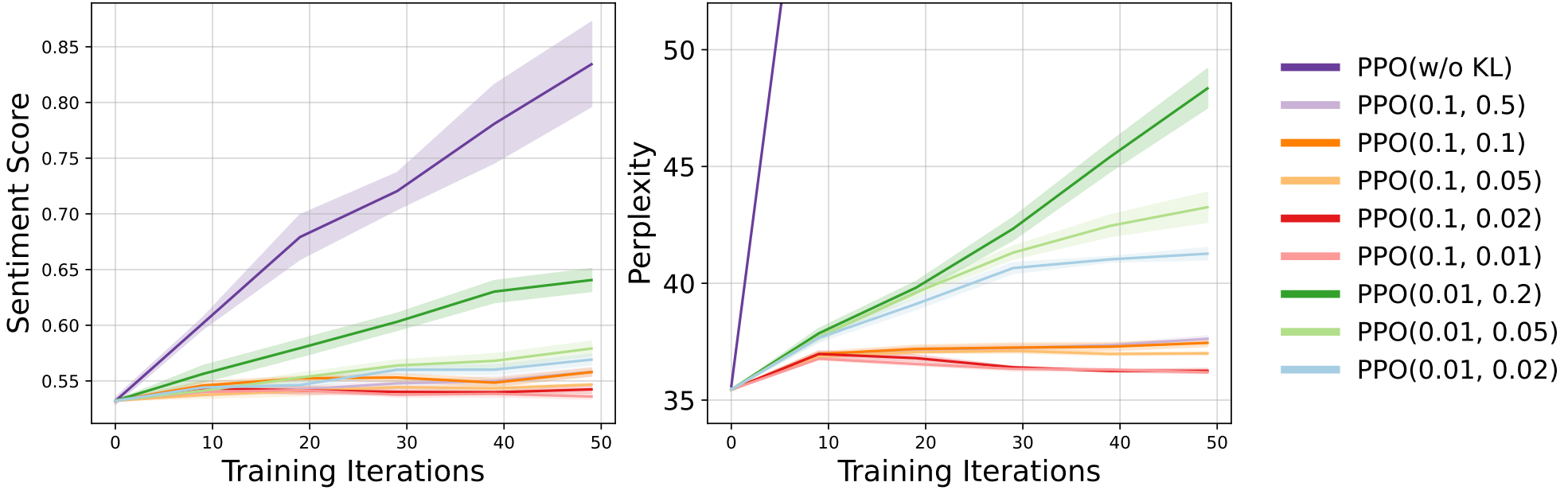

Sentiment Score
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
10
20
30
40
50
Training Iterations
Perplexity
50
45
40
35
0
10
20
30
40
50
Training Iterations
PPO(w/o KL)
PPO(0.1, 0.5)
PPO(0.1, 0.1)
PPO(0.1, 0.05)
PPO(0.1, 0.02)
PPO(0.1, 0.01)
PPO(0.01, 0.2)
PPO(0.01, 0.05)
PPO(0.01, 0.02)