強化学習のための大規模視覚言語モデルによるタスク進捗推定

吉村 亮哉[1], 太田 佳[1], 佐藤 育郎[1], 星野健太[1], 金崎 朝子[1], 川上 玲[1]

[1]東京科学大学

導入

背景

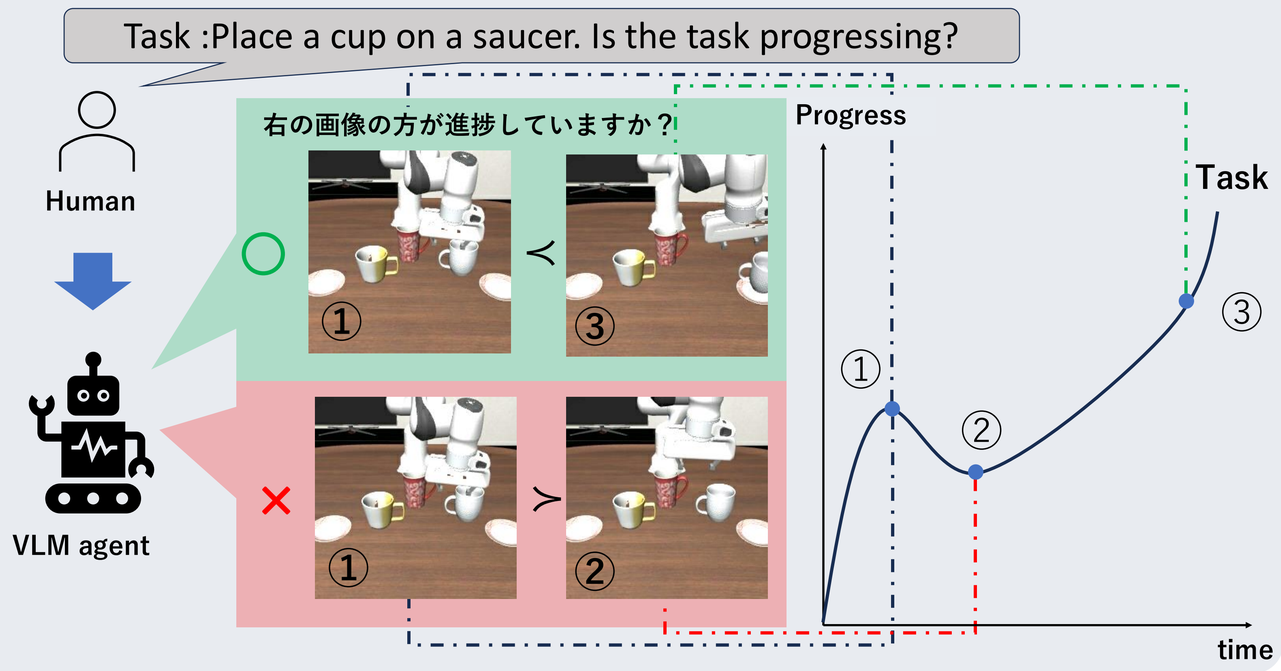
- ロボット操作の強化学習では，疎な報酬では探索が非効率
- 密な報酬の設計は専門知識と試行錯誤に大きく依存

目的

- タスクに汎用的な密な報酬の自動生成

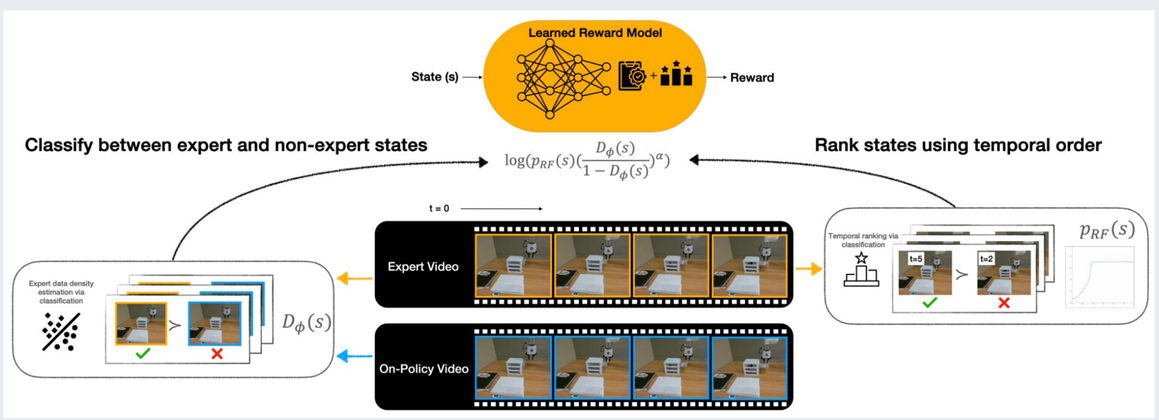
方針

- 軽量な大規模視覚言語モデル(VLM)をタスク進捗推定としてファインチューニング
- その出力をリアルタイム報酬として利用

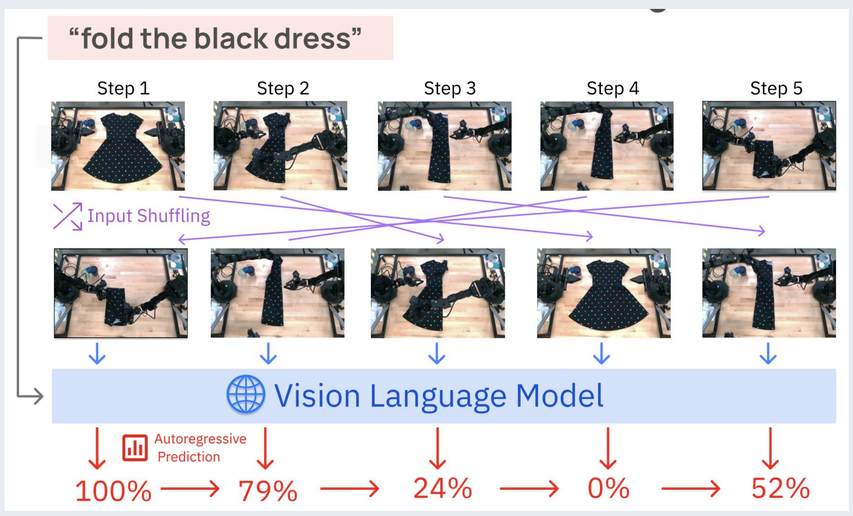


関連研究

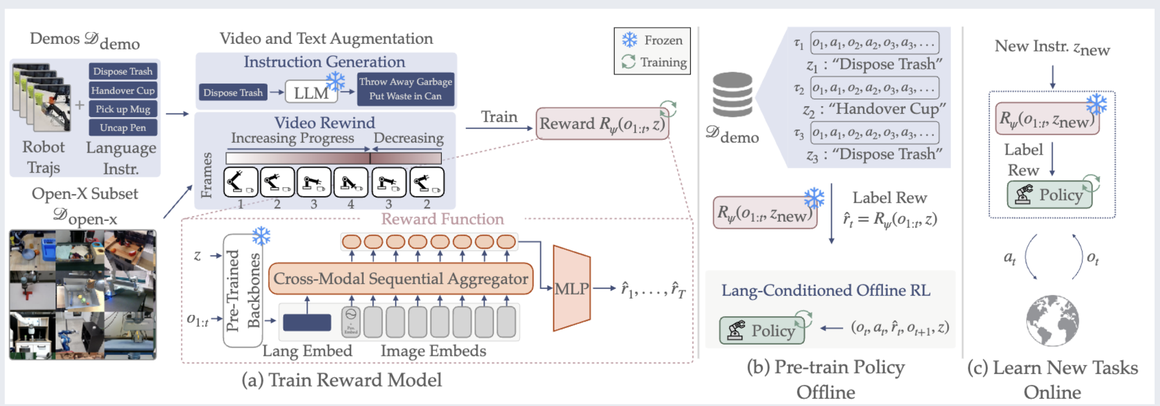
- Rank2Reward [1]
  - 映像の2フレームを比較して学習し，タスク進捗を推定



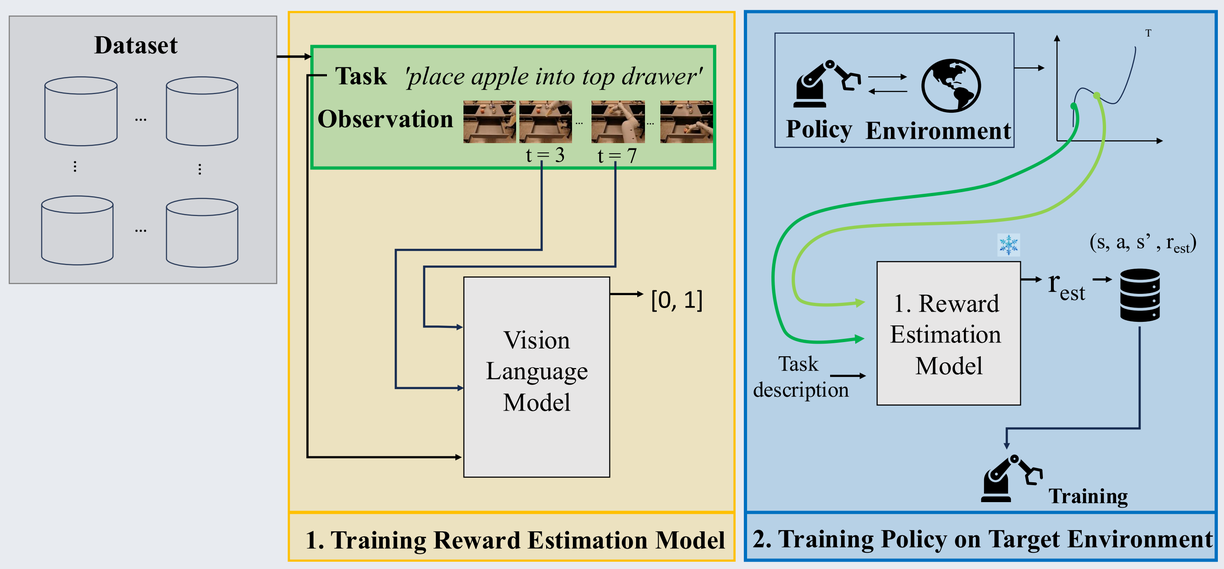
- Generative Value Learning [2]
  - シャッフルされた映像フレームのタスク進捗を大規模視覚言語モデルによりゼロショットで推定



- ReWiND [3]
  - 成功動画のみから成功例と失敗例をの軌跡を生成し，動画の進行度に応じた報酬モデルを学習



提案手法



ゼロショットのVLMはタスク進捗推定には不足
ロボットへの教示動画を用いた進捗推定を教師あり学習する手法[1]にならい，2枚の画像について進捗しているか否かの2値判定を学習する
学習した進捗推定モデルを報酬関数として評価

今後の展望

- モデルの精度向上
- 作成した進捗推定モデルを用いたシミュレーションと実世界での実装

参考文献

[1] Ma, Yecheng Jason, et al. “Vision Language Models are In-Context Value Learners.” ICLR2025

[2] Yang, Daniel, et al. “Rank2Reward: Learning Shaped Reward Functions from Passive Video.” ICRA2024

[3] Zhang, Jiahui, et al. "ReWiND: Language-Guided Rewards Teach Robot Policies without New Demonstrations.” RSS2025