SDRL: Interpretable and Data-efficient Deep Reinforcement Learning
Leveraging Symbolic Planning

Daoming Lyu¹, Fangkai Yang², Bo Liu¹, Steven Gustafson³
¹Auburn University, Auburn, AL; ²NVIDIA, Redmond, WA; ³Maana Inc., Bellevue, WA

**Problem**
- Sequential decision-making with long horizon action sequence and sparse reward suffers from:
  - Poor data efficiency,
  - Lack of interpretability.
- Challenge: Montezuma’s Revenge
  - The avatar: climbs down the ladder, jumps over a rotating skull, picks up a key (+100), goes back and uses the key to open the right door (+300).
  - Vanilla DQN achieves 0 score (Mnih et al., 2015).

**SDRL: Symbolic Deep Reinforcement Learning**
- **Goal:**
  - Symbolic planning drives learning, improving task-level interpretability.
  - DRL learns feasible subtasks, improving data-efficiency.
- **Task decomposition.**

**SDRL: Symbolic Planner**
- **Symbolic Planner:** high-level symbolic planning based on intrinsic goal.
  - Intrinsic goal: a linear constraint on plan quality $\geq quality(\Pi_t)$, where $\Pi_t$ is the plan at episode $t$.
  - Plan quality: a utility function that sums up the gain rewards of subtasks in a plan.
  - Mapping from symbolic transition to subtask.

**SDRL: Controller**
- **Controller:** low-level policy control with DRL.
  - Intrinsic reward: pseudo-reward crafted by the human.

**SDRL: Meta-Controller**
- **Meta-Controller:** subtask learning evaluation.
  - Extrinsic reward: a function about $\epsilon$ where $\epsilon$ is a criterion that measures the competence of the learned subpolicy for each subtask.
  - $\epsilon$: success ratio (in our case).
  - Learnable subtask and unlearnable subtask.

**Experimental Results**
- **Symbolic representation and predefined subtasks**
- **Final solution and learning curves**

**Reference**

**Conclusion**
- We present the SDRL framework, and it is the first work on integrating symbolic planning with DRL that achieves both task-level interpretability and data-efficiency for decision-making.
- Future work will investigate on the transferability, and integration with automatic option discovery.