Multi-Agent L-NovelD - Architecture and Language

Maxime Toquebiau

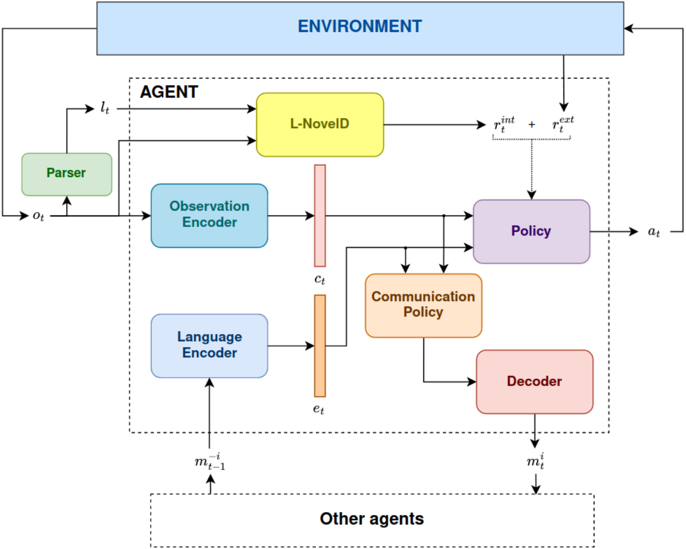
June 29th, 2022

Table of content

- Multi-Agent L-NovelD
  - Architecture
  - Modules
  - Contributions
  - Tasks
- Language conception
  - First iteration
  - Future iterations

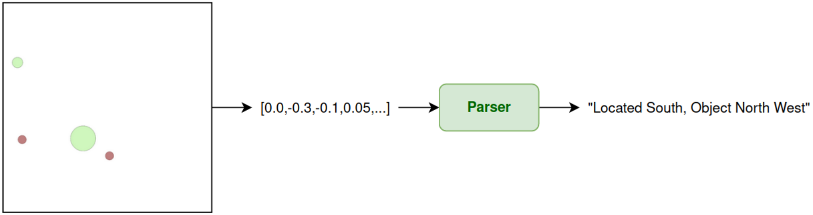
Multi-agent L-NovelD

Architecture



Multi-agent L-NovelD

Modules: Observation parser



[0.0,-0.3,-0.1,0.05,...]

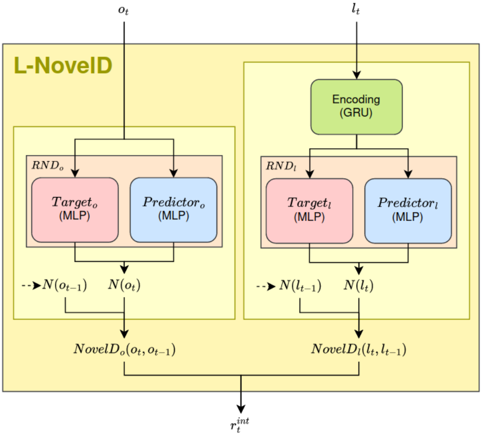
Parser

"Located South, Object North West"

- Rule-based
- Describes important elements of the observation

Multi-agent L-NovelD

Modules: L-NovelD



$$NovelD_l(l_t, l_{t-1}) = max(N(l_{t-1}) - \alpha N(l_t), 0).\mathbb{1}(N_e(l_t) = 1)$$

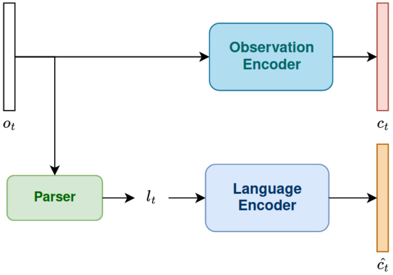
$$r_t^{int} = NovelD_o(o_t, o_{t-1}) + \lambda_l NovelD_l(l_t, l_{t-1})$$

Multi-agent L-NovelD

Modules: Language Encoder

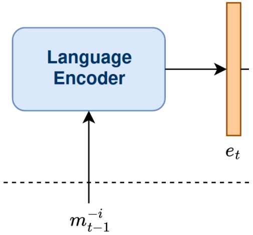
Training:

- Contrastive learning (Radford et al., 2021)[1]



Use:

- Encode messages from others agents

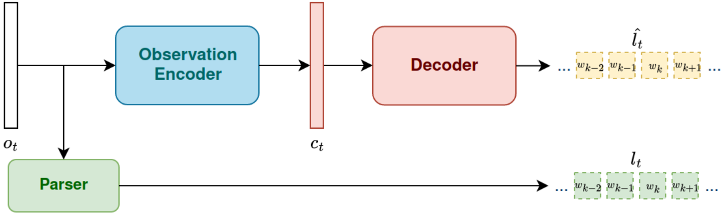


[1] Learning Transferable Visual Models From Natural Language Supervision, Radford et al., 2021

Multi-agent L-NovelD

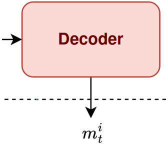
Modules: Language Decoder

Training: Observation captioning



o_t
Observation Encoder
c_t
Decoder
$\hat{l}_t$
... w_{k-2} w_{k-1} w_k w_{k+1} ...
Parser
l_t
... w_{k-2} w_{k-1} w_k w_{k+1} ...

Use: Generate messages



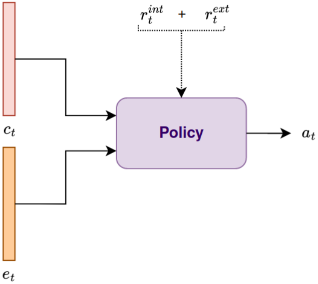
Multi-agent L-NovelD

Modules: Policy

Same policy as before:

- MADDPG
- QMIX
- ...

Trained with extrinsic reward from the environment **and** intrinsic reward from L-NovelD.

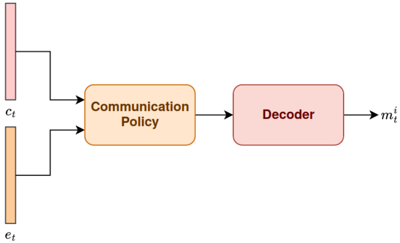


Multi-agent L-NovelD

Modules: Communication Policy

Goal: decide what to communicate, messages must be:

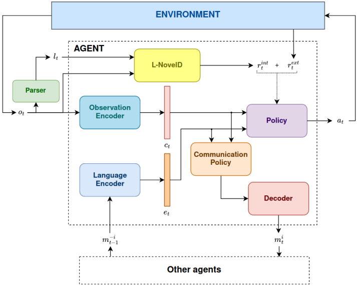
- **Valuable to other agents:** train to maximise the effect of messages on global reward (or on other agents' policy?)
- **Conform to reality:**
  - Must have some form of memory (e.g. recurrent neural networks)
  - Penalised when sending false information?



Multi-agent L-NovelD

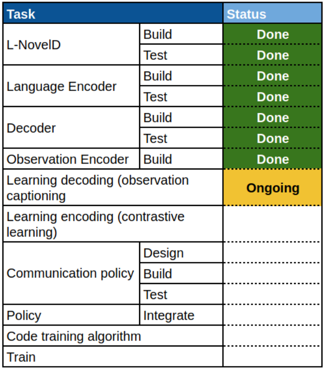
Contributions

- L-NovelD in a multi-agent environment
- Communication Policy
- Learn a language (encoding and decoding) and use it for communication



Multi-agent L-NovelD

Tasks



Task		Status
L-NovelD	Build	Done
	Test	Done
Language Encoder	Build	Done
	Test	Done
Decoder	Build	Done
	Test	Done
Observation Encoder	Build	Done
Learning decoding (observation captioning		Ongoing
Learning encoding (contrastive learning)		
Communication policy	Design	
	Build	
	Test	
Policy	Integrate	
Code training algorithm		
Train		

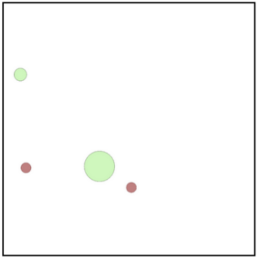
Language conception

First iteration

Vocabulary:

"Located", "Object", "Landmark", "North", "South", "East", "West", "Center", "Not"

Example:



"Located South, Object North West"

Issue with actions:
Actions are temporally extended ideas ⇒ not fitted to current model

More complex environment:
Add elements to show the advantage of language

- Colours
- Sizes
- Shapes

Questions?