Reward shaping & Language Augmented RL

Maxime Toquebiau

March 8th, 2022

Since last time

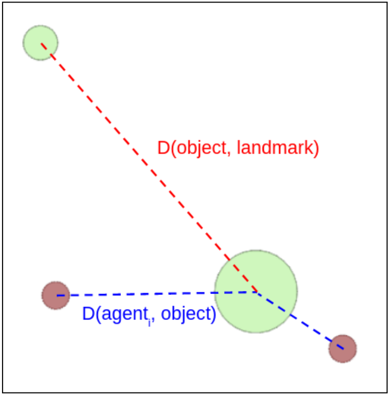
- Reward shaping
- Literature review: Language Augmented RL

Reward shaping

Previous reward

$$
\begin{aligned}
R^t_{dist\_reward} = & -D(object, landmark)^2 \\
& - \frac{1}{n_{agent}} \sum_{i=1}^{n_{agent}} D(agent_i, object) \\
& + \rho,
\end{aligned}
$$

with ρ the collision penalty, fixed at -10.



Reward shaping

New reward

- Sparse reward: big positive reward for success

  $\Rightarrow R_{sparse}(s_{t+1}) = \mathbb{1}_{success}(s_{t+1}) \times 50$

- $+$ penalty for every steps:

  $\Rightarrow R_{step}(s_{t+1}) = -0,1$

- $+$ shaping reward (Ng et al., 1999)[1]

  $\Rightarrow R_{shaped}(s_{t+1}) = \sigma(D_{obj,lm}(s_t) - D_{obj,lm}(s_{t+1}))$, with

$$\sigma = \begin{cases} 100, & \text{if } R_{shaped} > 0, \\ 10, & \text{if } R_{shaped} < 0, \end{cases}$$

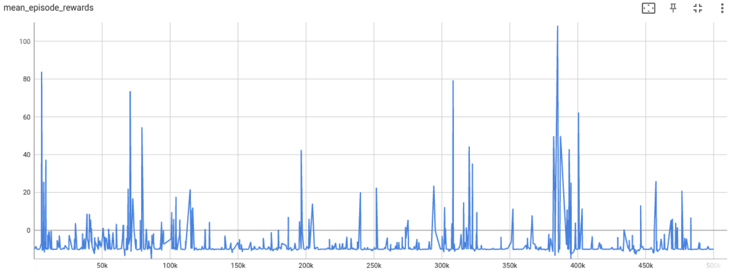
$$\Rightarrow R_{tot} = R_{sparse} + R_{step} + R_{shaped}$$

[1] Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping, Ng et al., 1999

Reward shaping

Results

MADDPG



Often doesn't even push the object.

Can't learn a good strategy ?

Suffer from random initial positions ?

Our goal:

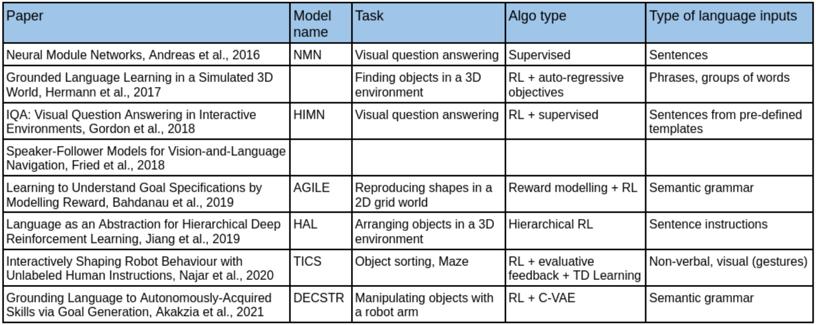
Teach a language to agents to give them:

- A **developmental tool**, to understand their environment, using language for generalising acquired knowledge.
- A **social tool**, to share information, to coordinate, to interact with humans.

Vygotsky: The developmental and social aspect of language develop concurrently.

Language-Augmented RL

Literature review



Paper	Model name	Task	Algo type	Type of language inputs
Neural Module Networks, Andreas et al., 2016	NMN	Visual question answering	Supervised	Sentences
Grounded Language Learning in a Simulated 3D World, Hermann et al., 2017		Finding objects in a 3D environment	RL + auto-regressive objectives	Phrases, groups of words
IQA: Visual Question Answering in Interactive Environments, Gordon et al., 2018	HIMN	Visual question answering	RL + supervised	Sentences from pre-defined templates
Speaker-Follower Models for Vision-and-Language Navigation, Fried et al., 2018				
Learning to Understand Goal Specifications by Modelling Reward, Bahdanau et al., 2019	AGILE	Reproducing shapes in a 2D grid world	Reward modelling + RL	Semantic grammar
Language as an Abstraction for Hierarchical Deep Reinforcement Learning, Jiang et al., 2019	HAL	Arranging objects in a 3D environment	Hierarchical RL	Sentence instructions
Interactively Shaping Robot Behaviour with Unlabeled Human Instructions, Najar et al., 2020	TICS	Object sorting, Maze	RL + evaluative feedback + TD Learning	Non-verbal, visual (gestures)
Grounding Language to Autonomously-Acquired Skills via Goal Generation, Akakzia et al., 2021	DECSTR	Manipulating objects with a robot arm	RL + C-VAE	Semantic grammar

How is language used to augment RL ?

How to interpret language ?

What language is learnt ?

Language-Augmented RL

Literature review

How is language used to augment RL ?

- **Goal description**
  - Hermann et al., 2017: phrase describing an object to pick
  - Bahdanau et al., 2019: semantic representation of goal state
  - Akakzia et al., 2021: semantic representation of goal state
- **Instruction following**
  - Jiang et al., 2019: high-level policy chooses instructions to follow
- **Structuring the model**
  - Andreas et al., 2016: modules with complementary roles
- **Induction of reward from language**
- **Text in the action or observation space**
- **Transfer from domain-specific textual resource**

Language-Augmented RL

Literature review

How to interpret language ?

- **Learning to encode language based on reward**
  - Jiang et al., 2019: GRU encodes instructions, parameters learnt with DQN
  - Gordon et al., 2018: LSTM + A3C
- **Decoupling language learning and policy learning**
  - Bahdanau et al., 2019: learn a reward model from goal states with supervised learning
  - Akakzia et al., 2021: learn to map sentence to semantic goal configuration with a VAE
- **Adding auxiliary objectives**
  - Hermann et al., 2017: unsupervised learning added to RL to help learning the language

Language-Augmented RL

Literature review

What language can be learnt ?

- **Semantic representations**
  - Bahdanau et al., 2019: e.g. “NorthFrom(Color(‘red’, Shape(‘circle’, SCENE)), Color(‘blue’, Shape(‘square’, SCENE)))”
- **Sentences, phrases (with pre-defined templates)**
  - Hermann et al., 2017
  - Gordon et al., 2018
  - Jiang et al., 2019
  - Akakzia et al., 2021
- **Non-verbal**
  - Najar et al., 2019: gestures

Percs of using language:

- Faster convergence
- Generalisation
- Curriculum design (Hermann et al., 2017)
- Human in the loop (Najar et al., 2019)

Issues with previous papers:

- Often rely on discrete states and actions
- Inflexible use of language
  - Pre-defined structures of sentences
  - Semantic representations ⇒ hard to understand/use by humans

Next steps

Reward:

- Continue exploring ideas
- Fixing bugs ?

Literature Review:

- Language-Augmented RL
- Hierarchical RL
- Communication with pre-defined language

Thank you!