

1711.10314 - Crossmodal Attentive Skill Learner

- **Yunqiu Xu**
 - NIPS 2017 HRL Workshop
 - Related reading:
 - A2OC, a prior work : [Harb et al, 2017. When waiting is not an option: Learning options with a deliberation cost](#)
 - Source code : <https://github.com/shayegano/CASL>
-

1. Introduction

- Challenges: similar to other HRL, this paper tries to solve
 - Durative tasks (long term)
 - Sparse reward
- Related insights:
 - Temporal abstraction: enables exploitation of domain regularities to form sub-goals (options)
 - Options (sub-goals) :
 - Improve learning by mitigating scalability issues in long-duration missions
 - Reduce effective number of decision epochs
 - Attention:
 - Focus on the most related parts
 - Capture longer-term correlations in its encoded state
- Our aim: learn rich skills that attend to and exploit multi-sensor signals at given moments
- Our work: CASL
 - Based on **option framework** and **A2OC**
 - Crossmodal : use multi-sensor (audio and video)
 - Attention: anticipate and identify usefule latent features, filter irrelevant sensor

2. Background

- I omit the part of POMDP and focus on "Options"
- Option framework: Sutton et al, 1999
- Similar to previous reading, here "option" is "sub-task"
- An option $\omega \in \Omega$ consists of:
 - Initiation set $I \subseteq \mathcal{S}$
 - Intra-option policy $\pi_\omega : \mathcal{S} \rightarrow \mathcal{A}$, this is **sub-policy**
 - Termination condition $\beta_\omega : \mathcal{S} \rightarrow [0, 1]$
- Given a state, master policy π select an option (suitable initiation set), then its intra-option policy will be executed to reach terminate state of this subtask \rightarrow a new state for next iteration until final end
- **A2OC**: prior work, extend A3C to Option-Critic

- $Q_\Omega(s, \omega)$: Option value function for option $\omega \in \Omega$

$$Q_\Omega(s, \omega) = \sum_a \pi_\omega(a|s) \left((r(s, a) + \gamma \sum_{s'} T(s'|s, a) U(s', \omega)) \right)$$

- $U(s', \omega)$: option utility function

$$U(s', \omega) = (1 - \beta_\omega(s')) Q_\Omega(\omega, s') + \beta_\omega(s') (V_\Omega(s') - c)$$

- If $\beta_\omega(s') = 1$, sub-task ends $\rightarrow U(s', \omega) = V_\Omega(s') - c \rightarrow$ Master policy
- If $\beta_\omega(s') = 0$, still sub-task $\rightarrow U(s', \omega) = Q_\Omega(\omega, s')$
- c : deliberation cost, add penalty when options terminate \rightarrow **let options terminate less frequently**
- $V_\Omega(s')$: value function over options (master policy π_{Ω})

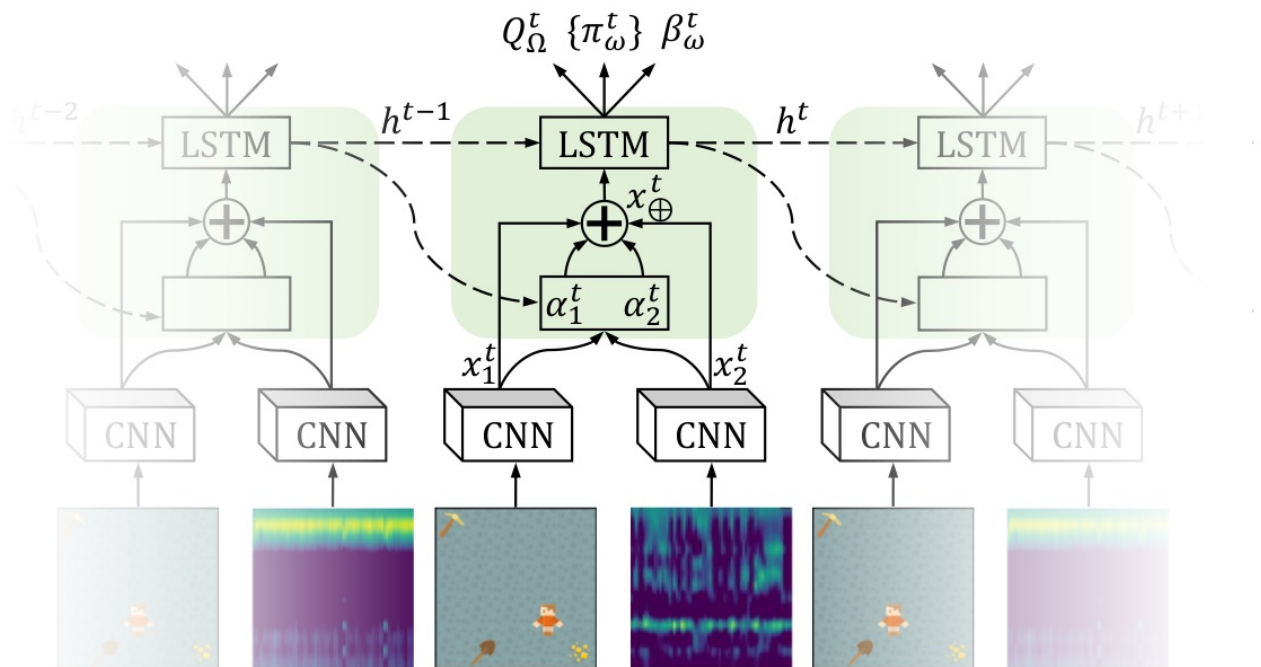
$$V_\Omega(s') = \sum_{\omega} \pi_\Omega(\omega|s') Q_\Omega(\omega, s')$$

3. Approach

3.1 Attentive Mechanisms

- Why we still need crossmodal attention?
 - Using crossmodal attention, agents combine internal beliefs with external stimuli
 - More effectively exploit multiple modes of input features for learning
 - Capture temporal crossmodal dependencies → faster and more proficient learning

3.2 Crossmodal Attentive Skill Learner



- M : the number of sensors
 - There are 2 sensors in this graph
 - Sensor 1 is image, sensor 2 is audio
- x_1^t : features extracted by CNN from sensor 1 at time t
- α_1^t : the relative importance for information from sensor 1 at time t

$$z^t = \tanh \left(\underbrace{\sum_{m=1}^M (W_m^T x_m^t + b_m)}_{\text{Exogeneous attention}} + \underbrace{W_h^T h^{t-1} + b_h}_{\text{Endogeneous attention}} \right) \quad (4)$$

$$\alpha^t = \text{softmax} (W_z^T z^t + b_z) \quad (5)$$

$$x_{\oplus} = \begin{cases} \sum_{m=1}^M \alpha_m^t x_m^t & \text{(Summed attention)} \\ [(\alpha_1^t x_1^t)^T, \dots, (\alpha_M^t x_M^t)^T]^T & \text{(Concatenated attention)} \end{cases} \quad (6)$$

- Eq (4):
 - Exogeneous attention: over sensory features x_m^t
 - Endogeneous attention: over LSTM hidden state h^{t-1}
- Eq (5):
 - Entropy regularization of attention outputs
 - Encourage exploration of crossmodal attention behaviors during training
 - 此处存疑: 这个 α 和之前的 "relative importance" 是一个东西么
- Eq (6):
 - Combine attended features $\alpha_m^t x_m^t$
 - Combine method: summed attention or concatenated attention
 - Then we can feed x_{\oplus}^t to LSTM

$$Q_{\Omega}(s, \omega) = W_{Q, \omega} h^t + b_{Q, \omega}, \quad (7)$$

$$\pi_{\omega}(a|s) = \text{softmax}(W_{\pi, \omega} h^t + b_{\pi, \omega}), \quad (8)$$

$$\beta_{\omega}(s) = \sigma(W_{\beta, \omega} h^t + b_{\beta, \omega}), \quad (9)$$

- LSTM captures temporal dependencies to estimate:
 - Option values: Eq (7)
 - Intra-option policies: Eq (8)
 - Termination conditions: Eq (9)

4. Experiment

- Learning tasks with inherent reward sparsity and transition noise
 - Door puzzle domain
 - 2D-Minecraft domain

- Arcade Learning Environment
- Evaluation:
 - Performance of CASL : learning rate and transfer learning
 - Understand relationships between attention and memory mechanisms
 - Crossmodal learning: modify ALE to support audio queries

4.1 Performance of CASL

- Attention improves LR, accelerates transfer

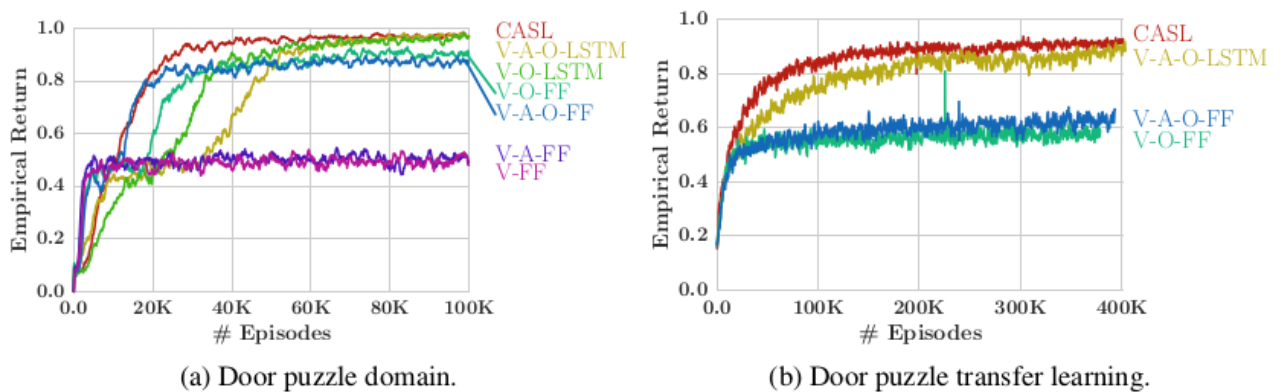


Figure 2: CASL improves learning rate compared to other networks. Abbreviations: **V**ideo, **A**udio, **O**ptions, **F**eedForward net, **L**STM net.

- Attention Necessary to Learn in Some Domains

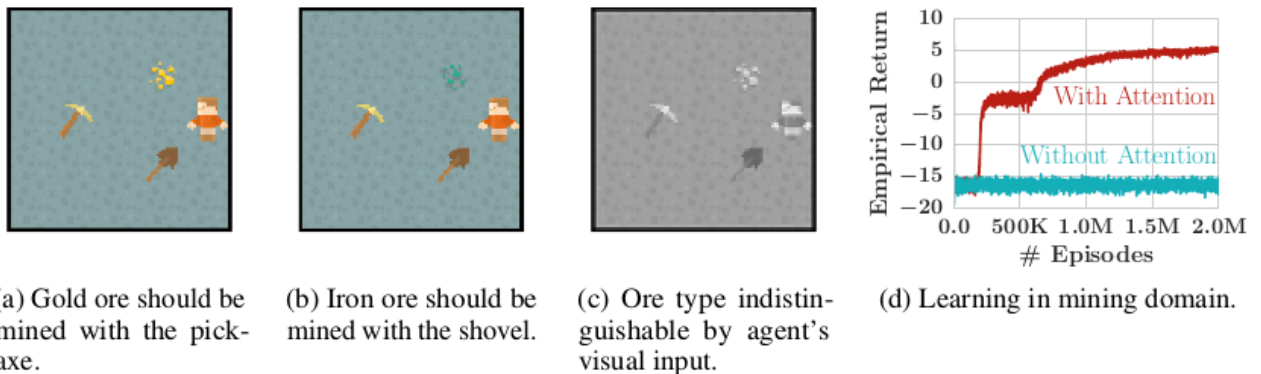
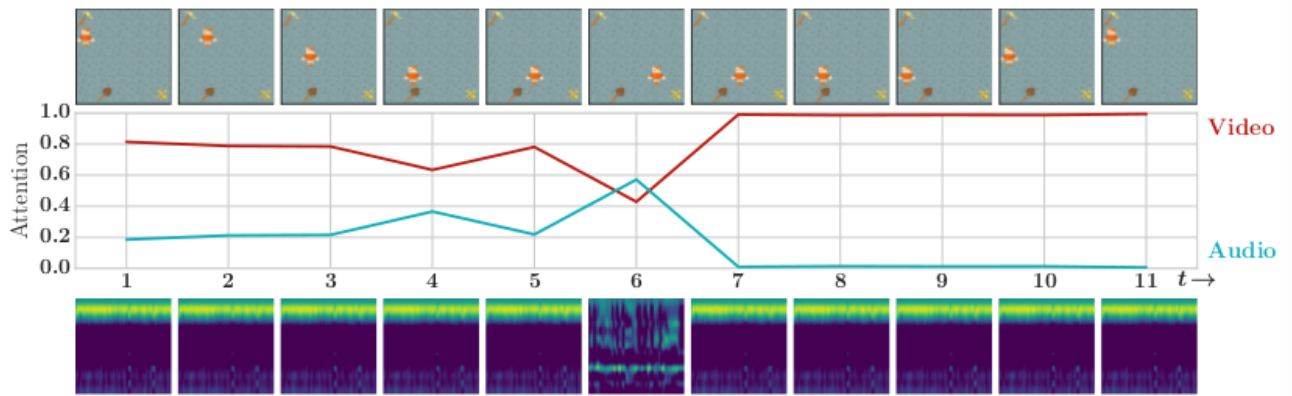
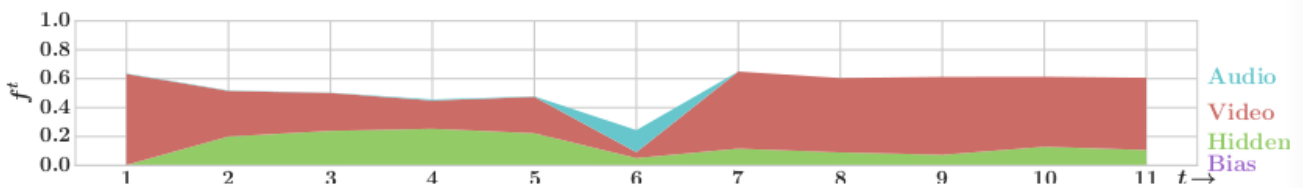


Figure 3: Mining domain. Ore type is indistinguishable by grayscale visual input to agent's network.

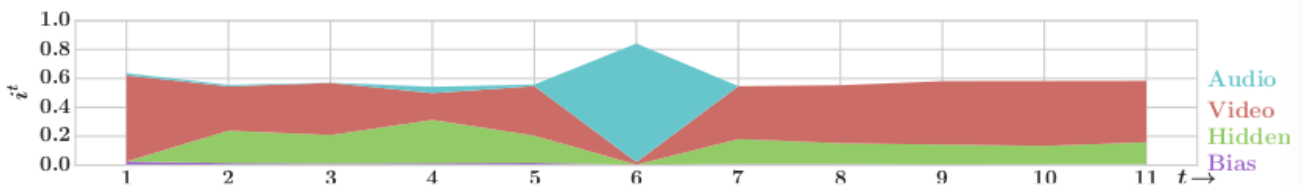
4.2 Interactions of Attention and Memory



(a) Agent anticipates salient audio features as it nears the ore, increasing audio attention until $t = 6$. Audio attention goes to 0 upon storage of ore indicator audio in the LSTM memory. Top and bottom rows show images and audio spectrogram sequences, respectively. Attention weights α^t plotted in center.



(b) Average LSTM forget gate activation throughout episode. Recall $f^t = 0$ corresponds to complete forgetting of the previous cell state element.



(c) Average LSTM input gate activation throughout episode. Recall $i^t = 1$ corresponds to complete throughput of the corresponding input element.

Figure 4: Interactions of crossmodal attention and LSTM memory. At $t = 6$, the attended audio input causes forget gate activation to drop, and the input gate activation to increase, indicating major overwriting of memory states. Relative contribution of audio to the forget and input activations drops to zero after the agent hears the necessary audio signal.

- Fig (4.a) : Before $t = 6$, the audio signal is nearly "non-useful" \rightarrow we need to check whether it's necessary to pay attention on it
- Fig (4.b) and Fig (4.c) : Overall activations for forget and input LSTM gates
- Result:
 - Before $t = 6$, the contribution of audio is zero (b/c), despite the attention in a is positive
 - When $t = 6$, forget gate drops and input gate increases \rightarrow overwriting of previous memory states with new information
 - Thus attended audio input is the key contributor
 - After $t = 6$, move attention back to video \rightarrow **listen to audio, but choose not**

to embed it into memory until the appropriate moment

4.3 ALE

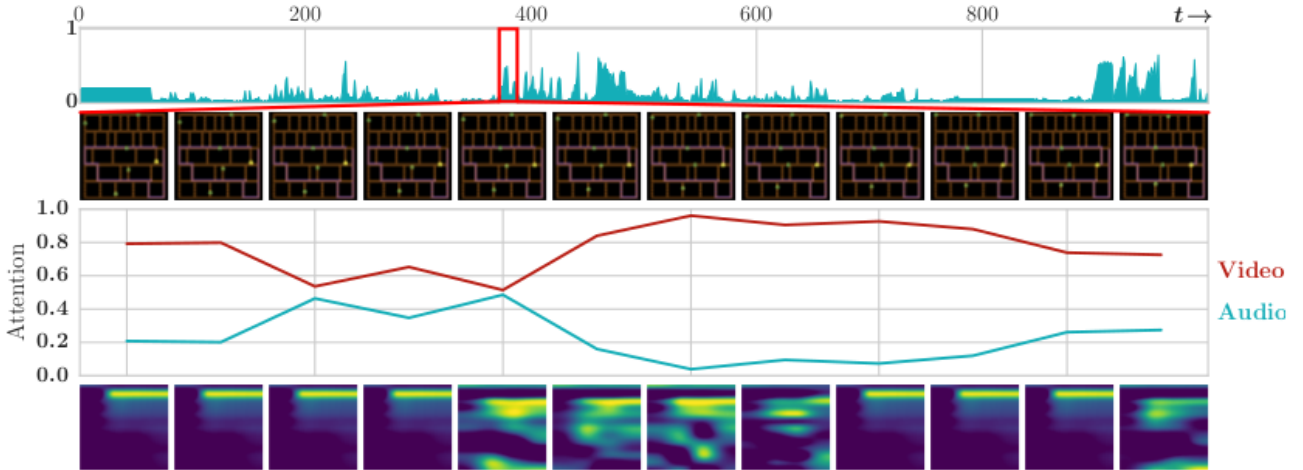


Figure 5: In Amidar, pathway vertices critical for avoiding enemies make an audible sound if not previously crossed. The agent anticipates and increases audio attention when near these vertices. Top row shows audio attention over 1000 frames, with audio/video/attention frames highlighted for zoomed-in region of interest.

Table 1: Preliminary results for learning in Atari 2600 game Amidar. Our crossmodal attention learner, even *without* option learning, achieves state-of-the-art score for non-hierarchical methods. We emphasize these are not direct comparisons due to our method leveraging additional sensory inputs, but mainly meant to highlight the performance benefits of crossmodal learning.

Algorithm	Hierarchical?	Sensory Inputs	Score
Mnih et al. [2015]	✗	Video	739.5
Mnih et al. [2016]	✗	Video	283.9
Babaeizadeh et al. [2017]	✗	Video	218
Ours (<i>without</i> options)	✗	Audio & Video	900
Harb et al. [2017]	✓	Video	880.0
Vezhnevets et al. [2017]	✓	Video	>2500

5. Summary

- CASL:
 - Integrate A2OC
 - HRL + multiple sensory inputs (video + audio)
- Feedback:
 - Interesting work, similar to human, we can make the agent to learn via both

video and audio input

- I need to check source code and prior work (A2OC) for more details