**Abstract learning: coherent representation of similar natured objects and actions of different games for better transfer learning**

**Abstract**

Reinforcement learning models can learn to play games just by interacting with game observations, actions and reward feedbacks. However, reinforcement learning models fails to transfer learning to a similar task if the visual style, action space and reward structure is different in that task. We humans on the other hand can learn and express diverse game objects and actions with general inter-related ideas. This paper discusses Abstract Learner approach that learns and represents similar natured game objects and actions coherently and passes it to reinforcement learner. Two theoretical model is discussed in this paper - a simple unsupervised model that coherently represents similar objects of different games in terms of their normalized reward nature, and a supervised model that learns and represents state objects in terms of semantic embedding. The supposition is that, coherent representation of similar natured objects of different games can enable reinforcement learning models to learn transferable policies.

**Problem with the current approach of RL**

Reinforcement learning (RL) is one of the most promising approaches in machine learning. RL agents can learn to play a game just by interacting with the state observation, action space and reward feedback, without the need for any background knowledge of the game. And with enough trial and error, they can learn important features from game observation and figure out good action policies that maximizes reward. It can even surpass human capabilities with new approaches to traditional games. However, this unsupervised nature also makes transfer learning challenging for RL. Some exciting work on general RL agent includes Importance Weighted Actor-Learner Architectures – IMPALA (Espeholt et al., 2018), which can simultaneously solve 30 different tasks of DMLab-30. Meta learning approach like Meta Learning Shared Hierarchies (Frans, Ho, Chen, Abbeel, & Schulman, 2017) aims to learn not just a particular task, but a range of related task by learning a high-level policy that generalizes to the common features. However, these approaches fail to transfer learning to a similar challenge, if the observation style, action space and reward structure is changed.

THE cops 
0.97 
1,349 YARDS 
Il?ll 
10 

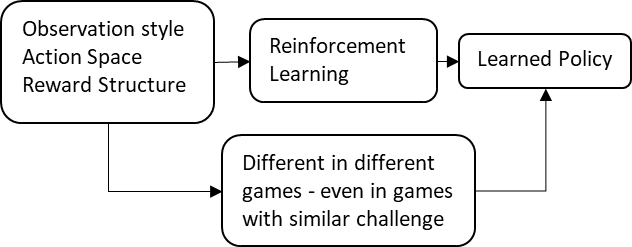


Figure 1: Reinforcement learning models struggles to transfer learnings between game with similar challenge but different observation style, action space and reward structure.

For learning features and features and action policies that are more transferable, an approach that learns to represent observation and action space of similar challenges coherently to the RL agent is required.

**How humans learn and express similar concepts coherently**

Humans are really good at understanding similarities of concepts of games, even if the concepts have visually different representation. We can understand and express continuously varying concepts of games with varying relation among a general source of discreet inter-related ideas represented by words. We learn words from our childhood, and we learn to express our complex and varying environment with varying inter-relation of words. Location of these ideas that words represents can be traced in our brain (Huth, De Heer, Griffiths, Theunissen, & Gallant, 2016).

For example, if we want to express the concept of a specific piece of paper, we might use inter-relationship of few discrete tangible ideas to express the paper – e.g. one A4 size cartridge paper; the concept might incorporate intangible ideas – e.g. ‘important’ information in the paper; the concept can even represent fictional idea – e.g. money (paper with fictitious value). Among thousands of different concepts, we can easily think of concepts that are closely related to paper and their degree of relatedness – e.g. pen, book etc. We can use these ideas to explain the nature of the specific paper to others, so that they get a general idea about the paper in question. Learning and expressing concepts in terms of general ideas that are shared among humans who speak the same language makes the learning and expression general and transferable.

Similarly, we understand games by understanding relationship of different ideas that the game objects represent. Even if the visual style, action space and reward structure is different in two games with similar challenge concepts, we can associate similar ideas with each of the games and transfer skills learned in one game to the other and vice versa.



Figure 2: Montezuma's Revenge.

For example, if we consider the game Montezuma's Revenge, upon seeing an animated object, we might understand that the object fits inside variance of the tangible idea 'Key', which is associated with concepts like lock, door etc. And upon seeing a door in that game, we might get an intuitive feeling that the solution to this task is probably to take the key to the door. After interaction with the game, we will associate fictitious idea of death with 'skull-like' object, even though we know we don’t usually die in real life if we touch a skull. While playing, we learn skills like moving, solving puzzle etc., and these skills can be transferred in another game with similar challenge, even if the action space, visual style and reward structure is different in that game. And we can even transfer the learning by telling others that the solution to this game is to take the key to the door while avoiding the skull, as the ideas of key, door and skull is general among those who speak English.

**Abstract learning: model for coherent learning and representation of similar game objects**

This paper discusses ‘Abstract Learner’ (AL) that following a similar approach as humans. It learns to coherently represent different game objects and actions that with similar nature with relationships of discreet general ideas. AL can be applied on top of RL, so that RL can learn more general features and policies that can be transferred among those similar challenges.

Game environments often depicts objects with fictitious nature, thus only visual information won’t be enough to explain those objects. For example, the turtle like creature in Mario represents fictitious negative reward, which is not a general characteristic that we associate with turtles. Thus, in addition to visual sensing, abstract learning agent needs to interact with game object to learn their nature.

**Unsupervised abstract learner**

Here we consider a simple unsupervised model for abstract learning of two simple games. In one game, the agent has to move past an obstacle to a rewarding object while avoiding a negative rewarding object. And in the other game, the agent has to drive a car.

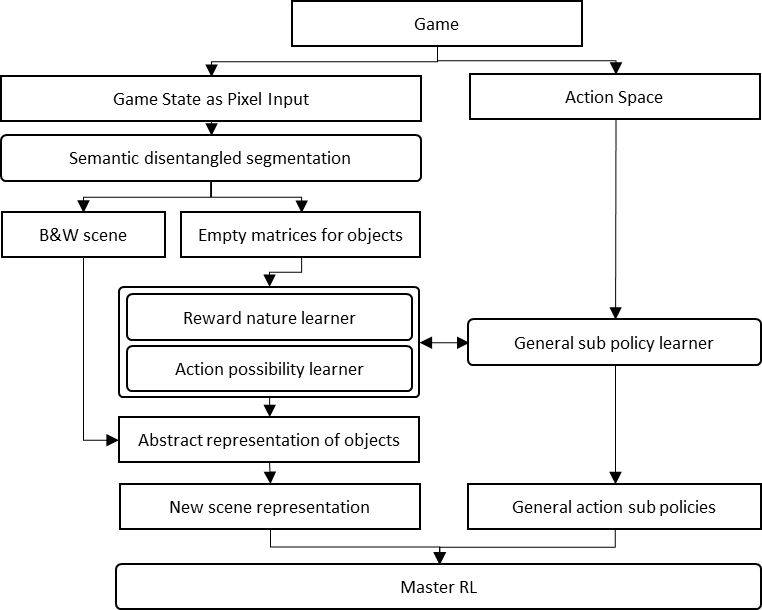


Figure 3: Simple abstract learner.

The model will have two parts, an abstract learner, and a core reinforcement learner. First, an unsupervised object disentanglement learner – e.g. MoNet (Burgess et al., 2019), will segment the objects in the state. The abstract learner will use just one abstract idea to express all the objects in the environment – reward nature of the objects of the environment. The model will also learn two general action sub policies for horizontal and vertical movement using actual game actions, as these two sub policies are sufficient for playing these two games. It'll interact with the objects and append the normalized reward value of each objects with the grayscale pixel values of those objects to create a new representation of the game. The original grayscale pixels of the games will be kept to reduce information loss. The core reinforcement learner will take in the new representation and action sub policies, and use that with reward feedback from the game to learn policies to play the game.

Second game

In the second game, the agent has to drive a car through road blocks to get reward. Getting out of the road will result in negative reward. In this case, the abstract learner will learn the same two master action sub policies that results in horizontal and vertical change in position, and reward nature of objects by interaction. The state object learner will append positive reward with road blocks and negative reward with outside field, and pass the new representation and action sub policies to the core reinforcement learner.

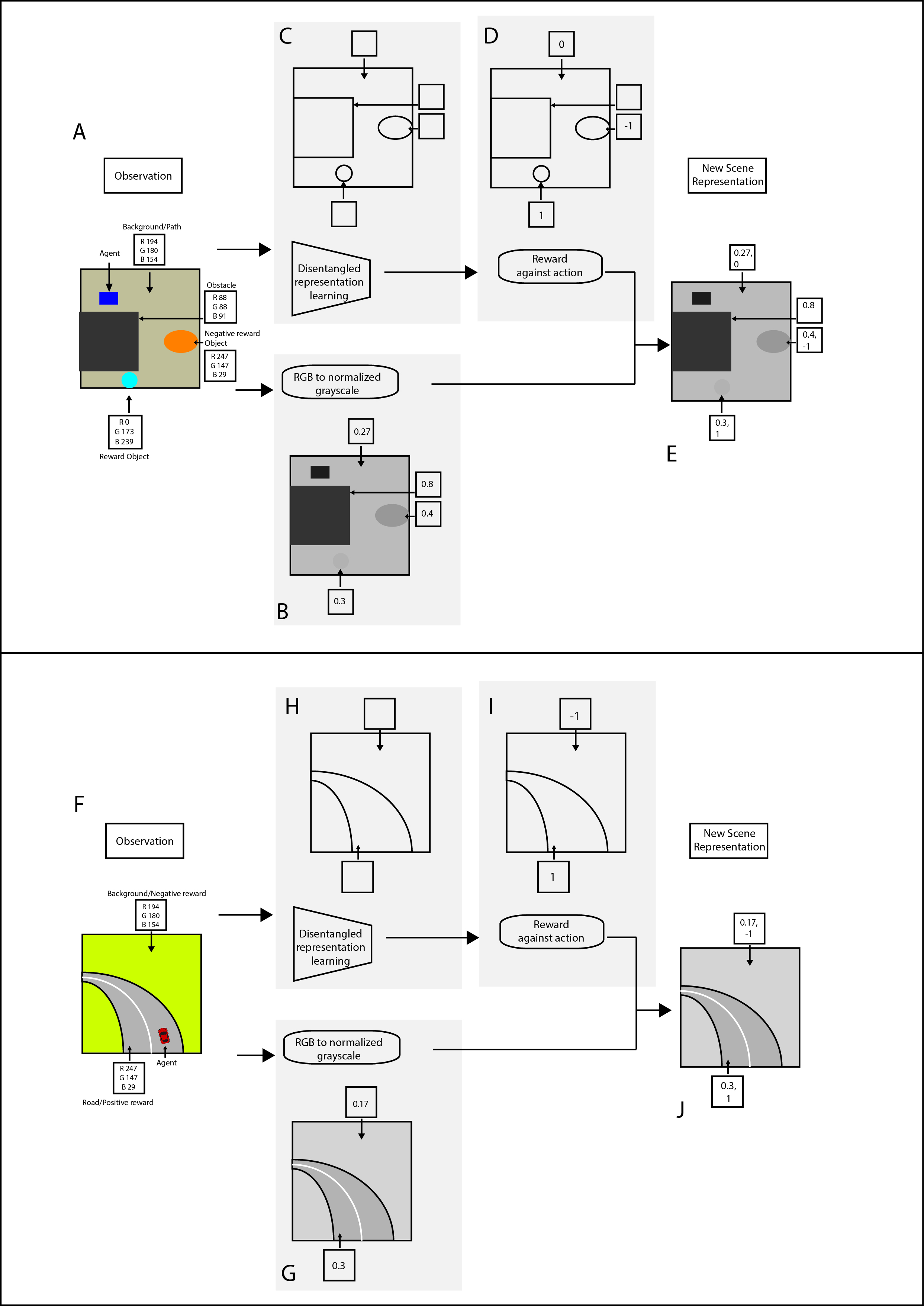


Figure 4: Intuitive illustration of simple abstract learner model.

The supposition here is that, if the reinforcement learner uses similar action sub policies to learn from observation that represents similar rewarding objects coherently across games, then the agent will be able to learn features that are transferable among the coherently represented games, even if the initial observation, reward structure and action space was different among those games.

However, in complex games with sparse reward, the intermediate objects have no reward feedback. Expressing them in terms of just rewards won’t preserve the contextual information of these objects, and thus the benefit of explaining state objects in terms of just normalized reward nature will become more and more trivial. In such case, supervised learner in addition to interaction learner will be necessary to learn and represent similar game objects.

**Supervised Abstract learner**

For a theoretical model for general learning, the main goal of the agent will be to maximize chance of its survival, and it will be able to do so by enriching its system of interrelated skills by solving more and more challenges.

The agent will have a memory system that contains interrelated continuous space of general ideas. In practice, we can use semantic embedding of inter-related tangible ideas expressed with mean and standard deviation. We'll call this portion the 'noun'. The agent will have visual sensing (e.g. using semantic segmentation) and interacting capacity for learning associated ideas about the environment it'll operate in. Upon sensing associated idea of an object from the environment, the agent will use those inter-related ideas to represent the object. The association of different object ideas (e.g. related concepts like lock and door in a frame) will help the agent define the challenge that the state is representing.

The agent will learn which sub policies or combination of the actual game actions produces results similar to general action results (e.g. horizontal, vertical position change). Using the new representations and sub policies, the agent will learn to solve the task with reinforcement learning; we’ll call these learnings the ‘verb’. Instead of initiating reinforcement learning with random actions, the agent will use idea reasoning – e.g. proximity of the idea to the concepts of opportunity or threat etc., and search which master action sub policies correlates with the state objects most (vertical change correlates with ladder etc.). Since the learnings is done against object system, learning to interact with related objects should help the agent learn interacting with a new object with less data.

The learnings will be saved in the memory in relation to the general state system of the game, so that the agent can reuse the learnings in case of a new game with a similar system of state objects. Since the learned policies are continuous in nature, the more game the agent plays, the more comprehensive its learnings will be, and the capacity to transfer learnings in new games will increase.

**Application of the model**

This paper has not discussed practical application and results of this method, and such experiment is yet to be executed. Rather is introduces intuitive discussion about possible approach.

There are some limitations that the outlined model fails to address, like representing intangible ideas of any object in a game.

In addition to promise of transfer learning, learning game features based on general source of explainable ideas can also help transfer learning among different agents that uses the same learning structure. The process of decision making by the agent may also be more conveniently understood by humans.

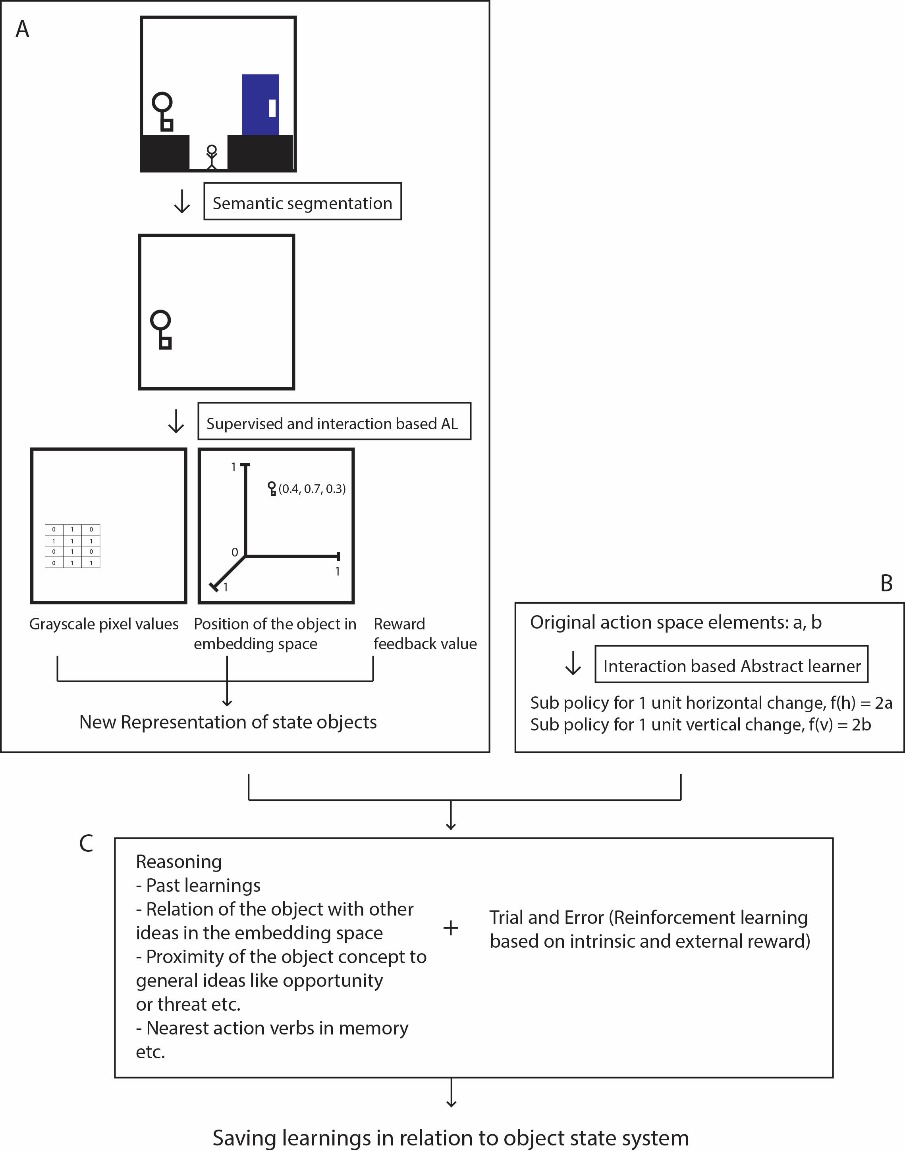


Figure 5: Supervised abstract learning model.

**References**

Burgess, C. P., Matthey, L., Watters, N., Kabra, R., Higgins, I., Botvinick, M., & Lerchner, A. (2019). *MONet: Unsupervised Scene Decomposition and Representation*. 1–22. Retrieved from http://arxiv.org/abs/1901.11390

Espeholt, L., Soyer, H., Munos, R., Simonyan, K., Mnih, V., Ward, T., … Kavukcuoglu, K. (2018). *IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures*. Retrieved from http://arxiv.org/abs/1802.01561

Frans, K., Ho, J., Chen, X., Abbeel, P., & Schulman, J. (2017). *Meta Learning Shared Hierarchies*. 1–11. Retrieved from http://arxiv.org/abs/1710.09767

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637