**Intuitions Regarding Transferable Predictive Knowledge**

Richard Feynman once explained most people will find ‘she went out, slipped on the ice, and broke her hip’ a reasonable answer to ‘why Aunt Minnie is in the hospital’ question. But for an entity without the knowledge of underlying causal relations like ‘when you break your hip you go to the hospital’, the answer might not make much sense.

Similarly, communication gets tricky for concepts which has no straightforward relation with personal motivation, or concepts which require understanding of many other underlying concepts that are unknown to the receiver. If I want to teach a student calculus from start, I need to make sure she has a sense that learning calculus can result in long-term benefits. I also need to understand her existing knowledge of math, since calculus requires understanding of numbers, slopes, rate of change, area under graph, solving equations etc. If she is motivated, and if I exactly know her existing level of knowledge, I can teach her required concepts.

Similarly, if we want to transfer answer of a question through predictive knowledge to an artificial agent, it is important that the agent has a similar notion of value regarding the question or task. Also, it is important to find out if the agent has necessary background causal knowledge regarding the answer or task – if the agent lacks background knowledge, it’ll require transfer of hierarchical predictions that wasn’t a direct part of the question asked in the first place.

For such transfer, the knowledge base of both the sender and receiver needs to be compatible – in a sense that if different artificial agents independently developed knowledge base by trying to solve a problem, they’ll develop similar and transferable knowledge. In case they solve interrelated problems (e.g. learning to play racing game and learning to drive autonomous vehicle), transferring knowledge learned from one task should help other. However, reinforcement learners learn feature as a function of given state-action-reward, and any change to those makes transfer hard, even if the nature of new task is similar. Furthermore, artificial agents’ knowledge can be represented using predictions about the future. Predictions can also be used to represent states (Littman et al., 2002). Predictive knowledge is autonomously verifiable, can be represented by general value functions, which can be learned by reinforcement learning approaches. But predictive knowledge has no public semantics, and is personal and subjective (Sutton et al., 2005; White, 2015).

The focus of this writeup is discussing design of internal environment of different agents (that computes state, generates rewards based on observation, executes actions given decisions etc. (Singh et al., 2010)) so that they can overcome some challenges of transferable knowledge.

**Problem 1: how to make knowledge relatable if different agents find arbitrary features in observation data stream to solve the task in hand:**

Firstly, we can consider similar distribution of environments and tasks that’ll be provided to the agent. The environment of a video game simulation differs vastly from the environment of real world, and thus the feature learned in one environment distribution might not help in another. Thus, we can consider similar real-world environment or real-world simulator for all agent to begin with.

Secondly, a significant human innovation is common language to express and think about reality and fiction. We think of similar ideas and share knowledge with others using common language. The location of certain ideas or words can be traced to regions of cerebral cortex (Huth et al., 2016).

We can consider similar approach of expressing observation with same reference. This can be done with human imitating supervised learner (e.g. semantic segmentation), which will detect objects or ideas of the environment and represent observations as graph of observed ideas and their relations. Thus, different artificial agents will have a similar low dimensional input of permutations of similar ideas given similar environments – which should lead to better and similar reasoning over those states with increasing training. However, supervised learning methods should be able to detect both concrete and abstract ideas, and also deconvolute each idea into their component ideas – like the idea of book can break down into paper, knowledge etc. Current semantic segmentation methods focus on 2D vision, but full spectrum 5 senses classifiers including 3D vision should help human imitating learning.

Relations of ideas in observed state can be learned with graph (Battaglia et al., 2018), and can also be used to solve RL tasks (Jiang et al., 2018; Madjiheurem & Toni, 2020).

Thirdly, the agents should have common action ideas to impact external environment, with possibilities of learning hierarchical actions. Options framework deals with learning higher level hierarchical decisions that execute many low-level decisions (Sutton et al., 2005). In case of humans, basic apparent low-level actions (use of hands and feet) can be thought as common. We abstract over or build complex new hierarchical actions over the similar basic actions – for example driving includes controlling the steering with hand and idea of movement. Thus, we have shared understanding of ways of impacting our environment. Common basic action ideas make the learned knowledge relatable for human.

All artificial agents with similar environment, similar supervised observation reference and same basic action ideas should allow the learned features of different agents to be compatible.

**Problem 2: how can artificial agents relate value of similar states if the extrinsic reward is arbitrary (e.g. applying different numeric reward for similar tasks):**

All human behaviors are for reducing primary deficit or drivers (survival and reproduction needs – hunger, sex, pain avoidance) or reducing learned secondary drivers – which is reinforced by predicting reduction of primary drivers. Animals get motivated to take action to restore balance of internal environment (e.g. have food if it’s hungry) (Skinner & Hull, 1944). Basis fitness function gives rise to primary, intrinsic and learned rewards. Like direct drivers, intrinsic rewards like curiosity are also inherent, but they don’t directly reduce primary drivers. However, on the long run, they help us gather experiences and skills that increases our chances of primary driver reduction across different scenarios. Intrinsic motivation attributes are more or less universal not only for humans, but also for other animals, which also indicates the goal of intrinsic motivation is ensuring survival of species. In fact, all activities that reduce primary drivers can be thought as intrinsically motivated, as all such activities indirectly increases chance of survival (Singh et al., 2010).

All human has similar goal of survival, and they act on similar environment with similar constraints and challenges, and they have similar basis actions to impact the environment. Thus, human value similar states similarly (states with food are good, states with harm are bad), which ultimately results in similar policy towards survival.

In order for similar valuation of many artificial agents, we can consider human mimicking approach for them. Likes humans, artificial agents can use a single question to value every state – probability of survival from that state. If there is an optimal value function that accurately predicts the change in chances of survival given any state, task (e.g. food acquisition) or hierarchical skill acquisition, all agents with that intrinsic value function will value related states similarly, which should allow transferable value function. If the agent had perfect causal relationship of the world, then just providing the goal of survival would suffice (Samuelson & Swinkels, 2006). But since the agent doesn't have such understanding to start with, we need to design reward functions that reward interim states like eating food, mating or exploration. Inverse reinforcement learning can be used to approximately estimate reward function of a subject from provided trajectories (Abbeel & Ng, 2004). It can be used to understand human imitating reward function for diverse scenarios, and apply that reward function to artificial agents.

If the internal environment of the agent gives rise to motivation to reduce primary driver, the agent can autonomously update value of world and take actions according to greedy policy to reach states that restores internal environment balance.

**Problem 3: how to ensure development similar causal knowledge structure:**

Humans don’t try to solve complex tasks like driving just after birth. We start with basic skills like movement and navigation and slowly learn more complex and hierarchical skills like driving. Even if a human play a racing game for the first time, given s/he knows settings of the game, s/he already has knowledge of movement, constrained time, importance of collision avoidance etc.

When an RL agent tries to solve a complex RL task from scratch, it tries to do so without any understanding of the parts of the problem. It does often allow the agent to find novel solutions to the task, but it also creates problems like sparse reward and requirement of many samples.

But if the agent is trained in a bottom up process – solving basic problems before complex ones, not only it should be able to learn good cognitive map and transfer low-level skills to solve the given racing game task more effectively, but it should also be able to solve related tasks like autonomous vehicle simulation with same low-level skills. In addition, a hierarchical problem might not seem to have any connection with basis goal of survival at first, but bottom up training should help the agent draw the connection.

TD networks can be used to build higher level predictions from lower level predictions, hierarchical options are composed of many routine low level decisions (Sutton et al., 2005).

**Problem 4: how should the agent search for knowledge of relatable ideas given observation:**

Using old experience to solve new task is discussed in RL literature (Niekerk et al., 2019; Saxe et al., 2017).

In a new situation, we usually get a sense of the value of the state from the relationship of the ideas of the state (e.g. availability of food, exam next day etc.). The value of states is often different for different people given internal environment balance (a thirsty person will value water more) or awareness of complex reward structure (one might not value education if s/he isn't sure about its long-term benefits). But generally, the of states that more directly reduces primary drivers is positively valued.

Similarly, given internal environment balance driven goal (e.g. hungry) and allowed actions, top down supervised breakdown of observation into finite ideas, and representing those ideas as graph should reduce variables and allow artificial agents to formulate value of state from past histories of similar state ideas with less calculation, and at the same time make the new history relevant for other agents that takes observation input with similar graph representation. We can compare this with ‘Goal directed learning’, where old learnings are used to solve new task.

Since learnings of the different agents are compatible because of their similar environment, observation reference and basis maximizing intrinsic reward function, we can consider a model-based approach where many agents will update a common model of world. This knowledge base can be stored in an external memory, and the agents will have access to it. Use of external memory in machine learning is addressed in literature (Graves et al., 2016).

As supervised semantic segmentation is used to predict ideas in the observation, learnings against such ideas can be searched in the predictive knowledge base. Model-based search and planning methods like Monte Carlo Tree Search can be applied to find related experience in the model or knowledge base, and the knowledge can be used as inductive bias for solving the new task. If the agent’s learnings are similarly transferable, it can be used to update the knowledge base.

**Type of approaches**

Depending on the internal environment design, we can think of human imitating or natural reinforcement learner. As described earlier, human imitating transferable knowledge can be possible if knowledge is gathered with reference to human imitating external and internal environment – reward function, basis actions and observation reference, learning process etc.

Human imitating environments is necessary where the artificial agents can act and learn. Designing such environment can be challenging. Artificial systems can learn from interactions in the real world (with similar impact and feedback on the environment as humans). In future, it might also be possible to directly access the histories of human brain to train artificial agents.

One of the motivations of the study of machine learning was to allow artificial systems to learn the rules for solving a problem instead of hand designing them ourselves. But human imitating intelligence will require hand designed reward function, environment, observation reference etc. Instead, we can think of new kind of problem solution skills that won’t require significant human input.

We can initiate agents that can learn transferable knowledge in a constrained environment that is not human imitating. Like an environment with increasing entropy, while the goal of artificial agents will be to minimize entropy with competition for constrained resources. Thus, the agents will have more reward if they can reduce more entropy with less resources. Human intelligence is nothing but the best approach to solve problem of human survival, and problem of artificial survival can give rise to different kind of intelligence.

For allowing transferable knowledge, the agents can have a common unsupervised representation learner. Unsupervised representation learning for a number of related tasks (Espeholt et al., 2018), and unsupervised object segmented observation input (Burgess et al., 2019; Greff et al., 2019; Watters et al., 2019) is discussed in RL literature. Unsupervised learning can allow NN to capture more accurate representation with increasing training.

We can combine RL and evolutionary methods – the agents will generate knowledge in their lifetime with RL methods, and variations of the most successful agents will be passed on to next generations. Upon generations of knowledge buildup, we can expect the agents to start figuring out novel approaches to solve given problems, which'll be a kind of intelligence, although not human imitating.

**Conclusion**

With technologies like 'Neuralink', we're aiming to build physical connection between human and artificial systems. One day it might be possible for artificial agents to learn knowledge that is directly transferable with humans. Magnus Carlsen has gathered years of experience to become world champion in chess, perhaps one day the learnings can be suited and transferred to another human or artificial agent instantly. At the same time, it might be possible to suit and transfer objective approaching knowledge of artificial agents directly to humans – like transferring Alpha Zero's chess expertise directly to Carlsen. Agent-Human-Agent transferable knowledge can allow both humans and artificial agents to instantly build on top of each other’s knowledge to solve harder problems. That might lead to more and more objective approximation of causal relations of reality.

At the same time, we can develop totally new kind of intelligence that tries to solve altogether different problems. It's not clear what'll be the use of such intelligence, but the opportunities demand exploration.

**References**

Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, *346*, 1–8. https://doi.org/10.1145/1015330.1015430

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., … Pascanu, R. (2018). *Relational inductive biases, deep learning, and graph networks*. 1–40. http://arxiv.org/abs/1806.01261

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

Espeholt, L., Soyer, H., Munos, R., Simonyan, K., Mnih, V., Ward, T., Yotam, B., Vlad, F., Tim, H., Dunning, I., Legg, S., & Kavukcuoglu, K. (2018). IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures. *35th International Conference on Machine Learning, ICML 2018*.

Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S. G., Grefenstette, E., Ramalho, T., Agapiou, J., Badia, A. P., Hermann, K. M., Zwols, Y., Ostrovski, G., Cain, A., King, H., Summerfield, C., Blunsom, P., Kavukcuoglu, K., & Hassabis, D. (2016). Hybrid computing using a neural network with dynamic external memory. *Nature*, *538*(7626), 471–476. https://doi.org/10.1038/nature20101

Greff, K., Lopez, R., Rishabh, K., Nick, K., Chris, W., Daniel, B., Matthey, L., Botvinick, M., & Lerchner, A. (2019). *Multi-Object Representation Learning with Iterative Variational Inference*.

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

Jiang, J., Dun, C., Huang, T., & Lu, Z. (2018). *Graph Convolutional Reinforcement Learning*. 1–13. http://arxiv.org/abs/1810.09202

Littman, M. L., Sutton, R. S., & Singh, S. (2002). Predictive representations of state. *Advances in Neural Information Processing Systems*.

Madjiheurem, S., & Toni, L. (2020). Representation learning on graphs: A reinforcement learning application. *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*.

Niekerk, B. Van, James, S., Earle, A., & Rosman, B. (2019). *Composing Value Functions in Reinforcement Learning*.

Samuelson, L., & Swinkels, J. (2006). Information, evolution and utility. *Theoretical Economics*, *1*(1), 119–142. https://econtheory.org/ojs/index.php/te/article/viewFile/20060119/452/16

Saxe, A. M., Earle, A. C., & Rosman, B. (2017). Hierarchy through composition with multitask LMDPS. *34th International Conference on Machine Learning, ICML 2017*, *6*, 4616–4629.

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70–82. https://doi.org/10.1109/TAMD.2010.2051031

Skinner, B. F., & Hull, C. L. (1944). Principles of Behavior. *The American Journal of Psychology*, *57*(2), 276. https://doi.org/10.2307/1416955

Sutton, R. S., Rafols, E. J., & Koop, A. (2005). *Temporal Abstraction in TD Networks*.

Watters, N., Matthey, L., Bosnjak, M., Burgess, C. P., & Lerchner, A. (2019). *COBRA: Data-Efficient Model-Based RL through Unsupervised Object Discovery and Curiosity-Driven Exploration*. 1–24. http://arxiv.org/abs/1905.09275

White, A. (2015). *Developing a predictive approach to knowledge*.