The preparation for the meeting with Prof. Alexandre on 03/02/2021

Organized by Jianyong
Version 1.1

February 3, 2021

Contents

1 Introduction

Artificial Intelligence (AI) has been built on the opposition between symbolic problem solving that should be addressed by explicit models of planning, and numerical learning that should be obtained by neural networks [1, 2]. But it is clear that in ecological conditions, our cognition has to mix both capabilities and this is nicely carried out by our brains. Similarly, behavior is often modeled with Reinforcement Learning techniques with two opposite approaches. In a Model-Based (MB) approach, an explicit model of the world is available for choosing a behavior from its anticipated consequences. In a Model-Free (MF) approach, the behavior is chosen from contingencies evaluated from past experiences and compiled in state and action values. Recent results rather report more general strategies to be more realistic, including the hybrid combinations of both [3], as proposed in the Dyna approach. In short, the key idea is to have a series of MF behavior, learnt in different contexts and to develop a MB approach to explicitly learn in which context to trigger which MF behavior, when to update them or to create new ones. In both cases, we have the same picture, with robust implicit modules learnt in different contexts and an explicit module managing them. We suggest this is a very interesting framework to study and model problem solving.

The research expertise of the Inria-CAS joint team will synergize to provide a unique leverage to address this important issue. On the Chinese side, connectionist models like deep neural networks are adapted to avoid so-called catastrophic forgetting and to facilitate context-based information processing [1] [4, 5]. This is done by a clever mechanism of weight modification to protect previously learned associations, and by a module learning to detect and reuse corresponding contexts to flexibly alter the Stimulus- Response association learned by the neural networks. On the French side, models in computational neuroscience explore the capacity of neuronal structures like the hippocampus to categorize contexts [6] and investigate the role of the prefrontal cortex [7], known to modulate behavioral activity depending on the context. We propose here to associate our experiences to develop a more general framework for adapting neural networks to problem solving, thus augmenting their usability in AI and the understanding of brain reasoning mechanisms [2].

2 The motivation

2.1 Objective 1: Specifying the general architecture.

Based on general architectures proposed in Reinforcement Learning and connectionist modelling (for example the Dyna algorithm, [8]) and in Cognitive Science and computational neuroscience (for example the Task Set model, [9]), proposing strategies to select, train and adapt elementary Stimulus-Response associations , one important goal will be to analyze the characteristics, bio-plausibility and performances of existing solutions and specify one architecture, suitable for the problem solving tasks considered here.

2.2 Objective 2: Implementing computing mechanisms.

These general architectures share some computing mechanisms, the principles of which can be studied independently to prepare their implementation. First, several Stimulus-Response associations

[1] see the model in detail

[2] 2 steps: 1: to learn elementary behaviors in diifferent contexts (Hippo) 2. to trigger adapted behavior (PFC)
sub-problems: 1. learning correspond to adapt and update a behavior or decide to train a new behavior 2. meta cognition: how to decide which behavior to trigger for optimality

must co-exist, to avoid catastrophic forgetting. This can be done by learning context-dependent associations [4, 5] or by defining attentional mechanisms for a adapted competition between several associations [10]. From a biological point of view, the ventral and dorsal parts of the lateral Prefrontal Cortex [11] have been reported to modify attentional activity in the sensory and motor cortex in that aim. Second, the ability to efficiently extract pertinent contextual information from complex, noisy environments to guide the choice/attention towards the proper Stimulus-Response association will be considered. In this regard, the joint team will examine the mechanism used by the hippocampus and explore how to apply related mechanisms to artificial neural networks. Third, analysis of performance must be exploited to learn and select the right strategies, as it has also been investigated in connectionist and bio-inspired solutions, though with less robust solutions for the moment. For all these mechanisms, it will be important to share and compare experiences in both teams to propose more efficient solutions.

2.3 Objective 3: Designing a task for evaluation.

In order not to remain at the conceptual level, it will be important to choose tasks to help specify and evaluate the developed models. Related to problem solving, the tasks must select ways to reach well-specified goals by learning adapted procedures and ways to organize them toward the goals [3]. This is consequently beyond classical identification and control and must include such dimensions as conceptualisation and organisation of behavior, thus providing original contributions in the domain of neural networks.

2.4 Objective 4: Sharing and disseminating.

One important objective of the associate team will be to train young researchers to these kinds of neural networks and to their exploitation in an AI framework. Reconciling the problem solving and learning sides of AI is a major aim in research today and the associate team, if successful, will gain a high visibility allowing for publications with high impact. We also plan to organize specialized workshops in Europe and in China.

3 The state of the art

3.1 Catastrophic forgetting

For achieving continual learning, the conventional neural network models suffer from catastrophic forgetting [4], that is, training a model with new tasks interferes with previously leaned knowledge and leads to significant decreases in the performance of previously learned tasks. Specifically, the OWM model [4] in training a network for new tasks, its weights can only be modified in the direction orthogonal to the subspace spanned by all previously learned inputs, which helps the network to

[3] task to implement ? important topic: Problem solving: to be define
not only exploration but search for a goal in a problem space
game ?
agent in a maze

[4] Continual learning poses particular challenges for artificial neural networks due to the tendency for knowledge of the previously learned task(s) (e.g., task A) to be abruptly lost as information relevant to the current task (e.g., task B) is incorporated. This phenomenon, termed catastrophic forgetting, occurs specifically when the network is trained sequentially on multiple tasks because the weights in the network that are important for task A are changed to meet the objectives of task B.

find a weight configuration that can accomplish new tasks while ensuring the performance of learned tasks remains unchanged.

Although a system that can learn many different mapping rules in an online and sequential manner is highly desirable, such a system cannot accomplish context-dependent learning by itself. To achieve that, contextual information needs to interact with sensory information properly.The PFC receives sensory inputs as well as contextual information, which enables it to choose sensory features most relevant to the present task to guide action4,5,29. To mimic this architecture, authors added the CDP module before the OWM-trained classifier, which was fed with both sensory feature vectors and contextual information. The CDP module consists of an encoder submodule, which transforms contextual information to proper controlling signals, and a "rotator" submodule, which uses controlling signals to manipulate the processing of sensory inputs. The encoder submodule is trainable and learns in a continual way with the OWM. Mathematically, the context-dependent manipulation serves by rotating the sensory input space according to the contextual information, thereby changing the representation of sensory information without interfering with its content. The rotation of the input space allows for OWM to be applied for identical sensory inputs in different contexts.

For achieving artificial general intelligence requires agents are capable to learn and remember many different tasks, that is, the ability to learn consecutive tasks without forgetting how to perform previously trained tasks. While Kirkpatrick et al. [12] propose a model of elastic weight consolidation (EWC), which allows knowledge of previous task to be protected while learning new tasks, thereby avoid catastrophic forgetting. It does so by selectively decreasing the plasticity of weights and thus has certain parallels with neurobiological models of synaptic consolidation. In analytically tractable settings, the authors demonstrated that EWC can protect network weights from interference and thus increase the fraction of memories retained over plain gradient descent [5]. To the extent that tasks share structure, networks trained with EWC reuse shared components of the network. Also, this work showed that EWC can be effectively combined with deep neural networks to support continual learning in challenging reinforcement learning scenarios, such as Atari 2600 games.

This work shows an algorithm that supports continual learning—which takes inspiration from neurobiological models of synaptic consolidation —can be combined with deep neural networks to achieve successful performance in a range of challenging domains. In doing so, we demonstrate that current neurobiological theories concerning synaptic consolidation do indeed scale to large-scale learning systems. This provides prima facie evidence that these principles may be fundamental aspects of learning and memory in the brain.

3.2 Planning and adaptation strategies

Baldassarre et al. [13] [6] propose an architecture which learns action affordances and forward models based on intrinsic motivations and can later use the acquired knowledge to solve extrinsic tasks by decomposing them into sub-tasks, each solved with one-step planning. An affordance is operationalized as the agent's estimate of the probability of success of an action performed on a given object.

Doya et al. [14] [7] propose a modular reinforcement learning architecture for nonlinear, non-stationary control tasks, which is called multiple model-based reinforcement learning (MMRL). The basic idea is to decompose a complex task into multiple domains in space and time based on the

[5] see in more details

[6] contribution ? task decomposition see in more details link to intrinsic motivation

[7] contribution ? task decomposition see in more details link to intrinsic motivation

predictability of the environmental dynamics. The system is composed of multiple modules, each of which consists of a state prediction model and a reinforcement learning controller. The "responsibility signal," which is given by the softmax function of the prediction errors, is used to weight the outputs of multiple modules, as well as to gate the learning of the prediction models and the reinforcement learning controllers. The performance of MMRL was demonstrated for discrete case in a nonstationary hunting task in a grid world and for continuous case in a nonlinear, nonstationary control task of swinging up a pendulum with variable physical parameters.

3.3 Generalization

From these definitions, it can be seen that the individual's high adaptability to complex and dynamic environments is an important indicator of intelligence; it is also the consensus of scholars in different fields to assess the level of intelligence based on the ability to adapt to environmental changes.

The human brain is clearly a model of high environmental adaptability. People can not only continuously absorb new knowledge in the new environment, but also can flexibly adjust their behavior according to different environments.

4 Solutions

4.1 Successor representation

From the point of view of the computational efficiency, model-free algorithm estimates the value function from empirical data and does not need to traverse all states in the state space, so it is especially suitable for data-based such as neural networks Function fitter. But its disadvantage [8] is that once the distribution of the sampled data changes, or the environment changes, all the parameters learned before will be invalid, and even some subtle changes will cause a significant drop in performance, which is the so-called "catastrophic forgetting" question. Therefore, from the perspective of the flexibility of the algorithm, the model-free algorithm performs poorly.

The model-based algorithm is just the other way round. It is not as efficient as the model-free algorithm to save resources, but the model is more flexible. This is because the model-based algorithm has some prior knowledge of the model itself, so when the environment changes, the model can be modified accordingly from the model parameters, and the algorithm can still perform well. However, when the state space is relatively large, this type of algorithm will consume computing resources and even cannot be solved.

Is there exists an algorithm that can compromise between computational efficiency and flexibility? Successor Representation [9] was first proposed by Peter Dayan of MIT in 1993 [15] (Hereinafter we referred to as SR). Considering that the core of the TD algorithm is to estimate the value function from the current moment to the future, Dayan believes that this value is closely related to the similarity of subsequent states. If there is a good representation that can describe the transfer characteristics from the current state to a certain future state, the value function can be decomposed

[8] cf also pb of making prospective reasoning what-if analysis, paper DolanDayan 2013:
retrospective reasoning; in MF: decide from past learning
prospective reasoning: what-if analysis: in MB
with SR: using replays

[9] SR intermediate MB-MF + can make prospective ? +replay ?

into two parts, one part is this representation, and the other part describes the reward function. So he proposed the SR method, combining the advantages of TD learning and the flexibility of model-based algorithms, making this algorithm called the third type of reinforcement learning algorithm in addition to model-based and model-free [16, 17].

In order to study the physiological basis based on the characterization of SR, Momennejad et al. [18] published an article in the "Human Behavior" and did a lot of experiments on humans and rodents to prove that SR has a certain biological basis. In addition, Gershman [16] also conducted a detailed analysis of the computational logic and neurological basis of SR from the perspective of behavioral and neuroscience, and believed that SR as a good compromise, which has achieved computational efficiency and flexibility compared with model-based and model-free.

4.2 Meta-episodic reinforcement learning.

A key starting point for considering techniques for fast RL is to examine why initial methods for deep RL were in fact so slow. The first source of slowness in deep RL is the requirement for incremental parameter adjustment. A second source is weak inductive bias.

4.2.1 Episodic Deep RL: Fast Learning through Episodic Memory.

The episodic RL by keeping an explicit record of past events and use this record directly as a point of reference in making new decisions. When a new situation is encountered and a decision must be made concerning what action to take, the procedure is to compare an internal representation of the current situation with stored representations of past situations. The action chosen is then the one associated with the highest value, based on the outcomes of the past situations that are most similar to the present [10].

4.2.2 Meta-RL: Speeding up Deep RL by Learning to Learn.

The leveraging of past experience to accelerate new learning is referred to in machine learning as meta-learning. Recent work has shown how learning to learn can be leveraged to speed up learning in deep RL. Here, a recurrent neural network is trained on a series of interrelated RL tasks. The weights in the network are adjusted very slowly, so they can absorb what is common across tasks, but cannot change fast enough to support the solution of any single task. In this setting, something rather remarkable occurs [11]. The activity dynamics of the recurrent network come to implement their own separate RL algorithm, which "takes responsibility" for quickly solving each new task, based on knowledge accrued from past tasks. Effectively, one RL algorithm gives birth to another, and hence the moniker "meta-RL".

4.2.3 Episodic Meta-RL

In episodic meta-RL, meta-learning occurs within a recurrent neural network. However, superimposed on this is an episodic memory system, the role of which is to reinstate patterns of activity in the recurrent network. As in episodic deep RL, the episodic memory catalogues a set of past events, which can be queried based on the current context. However, rather than linking contexts with value

[10] +Neural Turing Machine (Graves) and Differentiable NN
[11] see in more details

estimates, episodic meta-RL links them with stored activity patterns from the recurrent network's internal or hidden units. These patterns are important because, through meta-RL, they come to summarize what the agent has learned from interacting with individual tasks.

In episodic meta-RL [12], when the agent encounters a situation that appears similar to one encountered in the past, it reinstates the hidden activations from the previous encounter, allowing previously learned information to immediately influence the current policy. In effect, episodic memory allows the system to recognize previously encountered tasks, retrieving stored solutions.

4.3 Tasks for evaluation (To be continued...)

The simulations except the scalable and effective by solving a set of classification tasks based on a hand-written digit dataset and learning games sequentially.

The performance of the algorithms was demonstrated for discrete case in a nonstationary hunting task in a grid world and for continuous case in a nonlinear, nonstationary control task of swinging up a pendulum with variable physical parameters [14].

5 The preparation for the INRIA-CAS meeting

5.1 The introduction of myself

Hi, My name is Jianyong and I'm very pleased to join the team of Mnemosyne at Inria as a post-doc from December 2020, and also I'm so honored to participate in this collaboration project. I obtained my PhD from the university of claude bernard lyon 1 last November, and my subject mainly focuses on designing biologically Inspired Cognitive Architectures and developmental learning. like we put an artificial agent in an unfamiliar environment and let it construct the perception of this environment from interactions with it without any prior knowledge, and also in the situations that the environment has been changed, the agent has capabilities of self-adaptation and flexibility to generate novel behaviors.

I obtained my master's degree from North China University of Electric Power with a major of computer science in 2017. At the same time I was a research intern in the group of CRIPAC in the institute of automation, Chinese Academy of Sciences, under the supervision of Professor Liang Wang and Associate Professor Yongzhen Huang from 2015 to 2017. my subject was focusing on image segmentation , gait recognition with deep learning algorithms and developed strategies for parallel computing with GPU clusters and cloud computing. So very nice to meet you.

5.2 The format of this meeting

The ways to attend this visio conference: attending separately via visio to this meeting.

5.3 Topics

- Catastrophic forgetting and the interruption for absorbing new information.
- Attentional process mechanism. [19]

[12] paper ? more details ?

- Reinforcement-based learning mechanism in humans. The flexibility of activation-based working memories depend critically on the presence of a *dynamic gating* mechanism, which controls the updating and maintaining of working memory representations. The updating properties of the gating mechanism are shaped by a reinforcement-based learning mechanism, which plays a critical role in the present model by triggering the updating of working memory representations when the categorization rules changes [10] . Specifically, by simulating the role of dopamine in regulating the frontal cortex in terms of an adaptive-critic mechanism, a rapid trial-and-error search process emerged. The different levels of representations, when combined with the trial-and-error control mechanism, provided a quick way of reconfiguring the categorization rule used by the network via top-down biasing of different posterior representations.

- Reinforcement learning in Neural Turing machines. Discrete Interfaces cannot be trained directly with standard backpropagation because they are not differentiable. It is most natural to learn to interact with discrete Interfaces using Reinforcement Learning methods. In this work, by considering an Input Tape and a Memory Tape interface with discrete access. The proposed model is to use the Reinforce algorithm to learn where to access the discrete interfaces, and to use the backpropagation algorithm to determine what to write to the memory and to the output, the model is called the RL–NTM. [20]

- Flexibility with working memories. The working memory refers to the process of maintaining information through persistent neural firing, which can be updated rapidly by simple changing the activation state of a set of neurons. While more long-term memories encoded in connection weights between neurons require structural changes to update, which could be much slower (but more enduring). Therefore, activation-based memories support more flexible processing in the sense that a variety of different strategies, rules, or goals can be quickly activated and de-activated [10].

5.4 After the meeting

Should have a document to record the memo of this meeting, attached with the timeline of the progresses in this collaboration. A detailed work schedule could be the best.

References

[1] Hubert L Dreyfus and Stuart E Dreyfus. Making a mind versus modelling the brain: artificial intelligence back at the branchpoint. In *Understanding the Artificial: On the future shape of artificial intelligence*, pages 33–54. Springer, 1991.

[2] Ron Sun and Frederic Alexandre. *Connectionist-symbolic integration: From unified to hybrid approaches*. Psychology Press, 2013.

[3] Ray J Dolan and Peter Dayan. Goals and habits in the brain. *Neuron*, 80(2):312–325, 2013.

[4] Guanxiong Zeng, Yang Chen, Bo Cui, and Shan Yu. Continual learning of context-dependent processing in neural networks. *Nature Machine Intelligence*, 1(8):364–372, 2019.

[5] Guanxiong Zeng, Xuhui Huang, Tianzi Jiang, and Shan Yu. Short-term synaptic plasticity expands the operational range of long-term synaptic changes in neural networks. *Neural Networks*, 118:140–147, 2019.

[6] Randa Kassab and Frédéric Alexandre. Pattern separation in the hippocampus: distinct circuits under different conditions. *Brain Structure and Function*, 223(6):2785–2808, 2018.

[7] Xavier Hinaut and Peter Ford Dominey. A three-layered model of primate prefrontal cortex encodes identity and abstract categorical structure of behavioral sequences. *Journal of Physiology-Paris*, 105(1-3):16–24, 2011.

[8] Richard S Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine learning proceedings 1990*, pages 216–224. Elsevier, 1990.

[9] Philippe Domenech and Etienne Koechlin. Executive control and decision-making in the prefrontal cortex. *Current opinion in behavioral sciences*, 1:101–106, 2015.

[10] Randall C O'Reilly, David C Noelle, Todd S Braver, and Jonathan D Cohen. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral cortex*, 12(3):246–257, 2002.

[11] Robert S Blumenfeld and Charan Ranganath. Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, 13(3):280–291, 2007.

[12] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[13] Gianluca Baldassarre, William Lord, Giovanni Granato, and Vieri Giuliano Santucci. An embodied agent learning affordances with intrinsic motivations and solving extrinsic tasks with attention and one-step planning. *Frontiers in neurorobotics*, 13:45, 2019.

[14] Kenji Doya, Kazuyuki Samejima, Ken-ichi Katagiri, and Mitsuo Kawato. Multiple model-based reinforcement learning. *Neural computation*, 14(6):1347–1369, 2002.

[15] Peter Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, 1993.

[16] Samuel J Gershman. The successor representation: its computational logic and neural substrates. *Journal of Neuroscience*, 38(33):7193–7200, 2018.

[17] Tejas D Kulkarni, Ardavan Saeedi, Simanta Gautam, and Samuel J Gershman. Deep successor reinforcement learning. *arXiv preprint arXiv:1606.02396*, 2016.

[18] Ida Momennejad, Evan M Russek, Jin H Cheong, Matthew M Botvinick, Nathaniel Douglass Daw, and Samuel J Gershman. The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9):680–692, 2017.

[19] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *arXiv preprint arXiv:1410.5401*, 2014.

[20] Wojciech Zaremba and Ilya Sutskever. Reinforcement learning neural turing machines-revised. *arXiv preprint arXiv:1505.00521*, 2015.

Appendices

A Attached appendices