



International Symposium on Artificial Intelligence and Brain Science

今号は2020年10月10-12日にオンライン開催した国際シンポジウムの特集号です。

Foreword by Karl Friston: Natural Intelligence - or Artificial Intelligence Meets Brain Science	2
Keynote lecture by Josh Tenenbaum: Building Machines That See, Think and Learn Like People	4
Discussion on Session 1 talks: Deep Learning and Reinforcement Learning	6
Discussion on Session 2 talks: World Model Learning and Inference	8
Discussion on Session 3 talks: Metacognition and Metalearning	11
Discussion on Session 4 talks: AI for Neuroscience and Neuromorphic Technologies	13
Discussion on Session 5 talks: Social Impact and Neuro-AI Ethics	16

シンポジウム詳細はwebサイトをご参照ください。

プログラム、アブストラクト: <http://www.brain-ai.jp/symposium2020>

講義、討論ムービー: <https://vimeo.com/showcase/aibs2020>



プログラム、アブストラクト



講義、討論ムービー

Natural intelligence – or artificial intelligence meets brain science

Professor, University College London **Karl Friston**

I am delighted to be invited to write this prefatory note. Especially after enjoying a superb symposium organised by Kenji Doya and his colleagues on artificial intelligence and brain science¹⁾. Many of us had reservations about how this international symposium would translate into an online (virtual) conference. We need not have worried—the entire experience was gracefully managed from Tokyo and offered a substantive and engaging opportunity for international exchange on all our favourite topics. In brief, this symposium set the standard—at least for me—for what an international conference in the post-COVID era should look like.

There were a host of challenging and trending themes discussed. The axis of discussion—between artificial intelligence and neuroscience—worked perfectly. This was due, in part, to the impressive line-up of presenters and panellists, featuring nearly all the ‘great and good’ in high-end machine learning and theoretical neurobiology. There were too many themes for me to cover here, so I will drill down on three crosscutting themes, some of which were explicit and others more implicit.

One theme—that dominated the earlier sessions—could be usefully summarised as **structure learning**: namely, getting the structure and form of models right. I am taking models to refer to a singular concept that emerged with many different names, ranging from ‘physics engines’ through to ‘internal’ models, ‘predictive’ models through to ‘forward’ models. In my world, these are read as generative models that generate (observable) consequences from (latent or hidden) causes. The central role of generative models should, perhaps, not have been surprising, given our mission to look at artificial intelligence from the perspective of neuroscience and vice versa. At the end of the day, this is

necessarily going to involve some kind of belief updating and implicit inference, even in the context of action and learning. So how did structure learning evince itself in the discussions? This was evident



at a number of levels. For example, ‘learning to grow a mind’ calls on the notion of growing or exploring a model space, with the right kind of representations and relational aspects. This is the sort of problem addressed in nonparametric Bayes and, one might argue, any form of meta-learning (e.g., meta reinforcement learning). In terms of pure inference, structure learning is a solved problem—solved via Bayesian model comparison based upon model evidence or its variational bounds. However, what is not solved is which models should be compared, or indeed how one should extend existing models.

One approach to this problem may lie in what constitutes a good model. One might consider the only tenable way of scoring a model is in terms of its marginal likelihood or model evidence. This is a potentially important move because evidence is equal to accuracy minus complexity. This means that good models are minimally complex, in the sense of minimum description or message length approaches in algorithmic complexity, through to Ockham's razor. The imperative to simplify models came out in many proposals that all rest upon removing certain conditional dependencies among representations of latent causes (e.g., neuronal codes through to the architectures of variational autoencoders). One obvious way of leveraging independence—in the service of minimising complexity—is to introduce hierarchical structures.

Hierarchical or deep architectures are defined simply by the absence of conditional dependencies and provide simple explanations for observed or training data. However, more interesting ways of minimising complexity also emerged. I was taken with the notion of modular codes and factorisation. Perhaps it should be of no surprise that factorized representation was a recurrent theme. Indeed, all of variational Bayes (that underwrites variational autoencoders and their evidence lower bound) rests upon a mean field approximation that is defined in terms of a factorized posterior. In some, many of the beautiful examples that were on offer, from both machine learning and neuroscience, spoke to modularity, factorisation or hierarchical structures that embody the simplest explanation for the data at hand.

A second major theme seemed to be the covert distinction between generative models with continuous versus discrete state spaces. This was clearly an issue for people in machine learning who have a commitment to gradient-based schemes (e.g., back propagation of errors). Much of the discussion of structure learning—in terms of ‘program learning’ or ‘algorithmic learning’ seem to fall into one of two camps. Either we were talking about continuous state and time models (of the sort found in deep learning) or discrete state space models (of the sort required for symbolic processing).

To my mind, this was a particularly important distinction; especially, if one could unpack many of the more categorical or symbolic (and quintessentially hierarchical) models of cognitive like processing in terms of discrete state space models. The inversion or learning of such models may, on one view, be easy; for example, with the use of explicit fixed-point solutions and softmax operators. On the other view, they present challenges for gradient descent schemes. One thing—which I suspect will be discussed more and more in the future—is that the problem of structure learning over the space of discrete models or architectures. This brings me to

the final theme namely, information geometries.

In the first presentation by one of several very notable discussants, Amari posed a challenging question. Is there one principle that underwrites the various flavours of artificial intelligence and neurobiology we were considering. He never mentioned information geometry; however, I wonder whether this was out of modesty or whether he was waiting for someone else to suggest information geometry as the framework within which to find a unifying principle.

I will close by suggesting that information geometry may be the right way to think about things. This feels tenable from several perspectives. First, if we are talking about sentient artefacts, then we are talking about people and machines that infer, learn to infer, and learn to learn. But inference is just movement on a statistical manifold. This means that any form of belief updating, learning and implicitly sentient behaviour has a natural calculus supplied by an information geometry. I am particularly committed to this—given that the free energy principle is effectively a variational principle of least action on a statistical manifold. From the point of view of gradient schemes, this is nothing more than finding paths of least action that minimise some functional that supervenes on and appropriate information geometry. I found myself musing on whether one could equip model spaces—over models—with an information geometry and thereby finesse the structure learning problem. Perhaps someone has already done this and, if they have, we will find out about it when we next meet.

Information geometries of a certain kind may also underwrite consciousness. I end with consciousness because this is how the meeting started with Yann LeCun’s suggestion that consciousness was the next step that artificial intelligence would need to consider – a step that is clearly licensed via its intimate relationship with the brain sciences.

1) <http://www.brain-ai.jp/symposium2020/>



International Symposium on Artificial Intelligence and Brain Science

Keynote lecture

Title : Building Machines That See, Think and Learn Like People

Speaker :



Josh Tenenbaum
Massachusetts Institute of Technology

Reporter : Yohei Oseki (The University of Tokyo)

Josh Tenenbaum (Massachusetts Institute of Technology) kicked off the international symposium with the keynote lecture titled “Building machines that see, think and learn like people”. This lecture updates the influential paper “Building machines that learn and think like people” published in Behavioral and Brain Sciences (Lake et al., 2017) on one hand, and the recent lecture “Reverse engineering common sense” given at Brains, Minds and Machines Summer Course 2020 on the other, with particular emphasis on brains. This report summarizes the main points of the lecture and several discussions raised after the lecture.

1. Keynote lecture

First, Tenenbaum pointed out the problem with current artificial intelligence (AI) that there is no “real” general-purpose AI which can flexibly solve multiple tasks such as object recognition and game playing. Specifically, current AIs à la deep learning have been restricted to pattern recognition and cannot be generalized to domains vastly different from training (e.g., Basketball scenes with various image resolutions), which humans can understand via

world models with essentially no re-training or fine-tuning.

In order to solve this problem, Tenenbaum proposed to reverse-engineer the so-called “Common Sense Core” of human intelligence. Specifically, there are two types of “Common Sense Core”, intuitive physics and intuitive psychology, both of which are the world models to simulate physical worlds of objects and mental worlds of agents, respectively. His team demonstrated this idea by “the game engine in your head”, which simulates physical and mental worlds just like game programmers. For the purpose of building this architecture, intuitive physics and intuitive psychology are computationally implemented with probabilistic programming languages (e.g., Pyro, TensorFlor, Probability, PyProb, Turing, Gen, etc.), that integrate symbolic structures and neural networks in the framework of Bayesian inference.

Then, recent experimental and computational results were summarized with special focus on “the intuitive physics engine” (Battaglia et al., 2013; Ullman et al., 2017) and “virtual tools” with Sample, Simulate, Update (SSUP) model (Allen et al., 2020). They explored how the intuitive physics engine was

realized in the brain and found their neural correlates in the premotor and parietal cortices, which overlap with action planning and tool-use regions of biological neural networks (Fischer et al., 2016; Schwettmann et al., 2019). In addition, intuitive physics was also modeled with artificial neural networks (Mrowca et al., 2018; Bear et al., 2020) which successfully invert graphics programs in primate vision (Yildirim et al., 2020).

Finally, he demonstrated how probabilistic programs are learned efficiently (Lake et al. 2015) with two specific learning algorithms. The first algorithm was “the child as hacker” (Rule et al., 2020), where children are assumed to “hack” (write code) through program induction and program synthesis. The second algorithm was DreamCoder (Ellis et al., 2020; Tian et al., 2020), where wake-sleep algorithms simultaneously learn symbolic structures and neural networks.

In summary, Tenenbaum concluded that humans (and machines, by extension) must have built-in starting states and learning procedures than we might have thought, by addressing two fundamental research questions in the cognitive science. For the first question “what is the starting state (inductive bias)?”, humans simulate physical and mental worlds with intuitive physics/psychology and “the game engine in your head”. For the second question “what are the learning procedures?”, humans learn world models via probabilistic programs and program induction/synthesis.

2. Discussion

Question (Kenji Doya): What made wake-sleep algorithms (i.e., the combination of symbolic structures and neural networks) computationally efficient, given that symbolic structures are claimed to be learned slowly in the recent lecture “Reverse engineering common sense”?

Answer: Traditional insights from programming language for the machine learning purpose were responsible for synthesizing symbolic structures and neural networks with the joint objective function under the hierarchical Bayesian framework.

Question (Wenliang Kevin Li): How abstract symbols are coded in the brain?

Answer: How natural languages and complex multi-object scene representation are coded in the brain will provide insights into this question.

Question (Nergis Tomen): How intuitive physics is learned?

Answer: Intuitive physics should be evolved phylogenetically via distributed evolutionary code writing, but may not be learned ontogenetically and, instead, innately determined with proto-intuitive physics, as experimentally demonstrated by Elizabeth Spelke (Harvard University) and Rebecca Saxe (Massachusetts Institute of Technology).

Question (Shriraj Sawant): How important is continual learning?

Answer: Simpler and easier tasks are learned early in development and “bootstrap” more complex and difficult tasks later as in curriculum learning.

The discussion continued on the Slack channel prepared for this session.

3. Conclusion

In conclusion, Tenenbaum provided the insightful big picture towards correspondence and fusion of artificial intelligence and brain science. In fact, Tenenbaum’s keynote lecture continued to be referred to by various talks and discussions throughout the international symposium. Higher-order cognitive functions such as intuitive psychology and natural language remain to be investigated in great detail and left as important frontiers for future research.

Discussion on Session 1 talks

Theme : Deep Learning and Reinforcement Learning

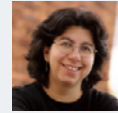
Speakers :



Yann LeCun
Facebook AI Research /
New York University



Yutaka Matsuo
The University of Tokyo



Doina Precup
McGill University



David Silver
DeepMind



Masashi Sugiyama
RIKEN AIP/
The University of Tokyo IRCN

Discussant : Shun-ichi Amari (RIKEN Center for Brain Science)

Chair : Jun Morimoto (ATR Brain Information Communication Research Laboratory Group)

Reporter : Tadashi Kozuno (Okinawa Institute of Science and Technology Graduate University)

Report

A discussion was on diverse but highly connected topics, which are the following:

1. A unified theory for deep learning and the mind;
2. The role of attention;
3. Components necessary for one-shot learning;
4. How to realize the higher cognitive functions, such as the reasoning and understanding of commonsense;
5. The involvement of language processing in higher cognitive functions.

In this report, I will explain each of them. Then, I conclude the report with my thoughts about these topics.

A Unified Theory of Deep Learning and the Mind

Prof. Amari raised a question on whether it is possible to establish a unified theory for deep learning. There are several theories of deep learning, but they are isolated and can capture only a small aspect of deep learning wherein they are specialized.

Dr. Silver argued that besides a unified theory for deep learning, we need a unified theory for the mind: what is an overall objective of the mind? For example, we divide a task into small subgoals and solve them one by one. When we look at a person trying to complete a subgoal, we can identify what they want to do. Nonetheless, we cannot understand what the

person is trying to archive in the end by solving subgoals. Therefore, to fully understand intelligence, we need to understand the overall objective of the mind, which he thinks is possible.

He also pointed out the importance of keeping in mind the overall objective instead of identifying and understanding each subgoal of a brain part. Each part of the brain is specialized in its task, but it exists to archive the overall objective.

On the contrary, attributing the success of AlphaGo to a combination of two distinct learning systems, Prof. Sejnowski argued that a unified theory for learning would be impossible since the brain is dozens of learning systems running parallelly and interacting with each other.

The Role of Attention

Prof. Wang asked why the brain can pay attention to only one thing and has a limited working memory capacity. Seemingly, such limitations are unnecessary.

A hypothesis by Prof. LeCun was the following: we need multiple world models specialized in different situations, but the size of the brain is limited; as a result, the brain has only one configurable world model. Since such a world model needs to be configured for a task at hand, we can pay attention or focus on only one thing at a time.

Components Necessary for One-shot Learning

Prof. Wang also asked what components are necessary for one-shot learning, besides a differentiable neural network.

Prof. LeCun answered as follows: in deep learning, we have neural network architectures similar to the hippocampus, such as memory networks, with which one-shot learning is possible. The next step for AI is to enable AI to reason by combining hippocampus-like and cortex-like neural networks.

How to Realize the Higher Cognitive Functions

Everyone agreed that it is important to understand the higher cognitive functions and implement them in (and with) deep learning models.

In particular, Prof. LeCun and Prof. Tenenbaum emphasized the importance of learning to reason with neural networks and gradient descent. As witnessed in the deep learning community, gradient descent is a scalable and powerful tool. However, a question is how to approximate the symbolic reasoning by neural networks; the symbolic reasoning is inherently discrete, whereas neural networks can output only continuous values.

The Involvement of Language Processing in Higher Cognitive Functions

Prof. Sejnowski referred to the following fact as a paradox: previously, symbolic AI was a dominant paradigm, but it was superseded by neural networks or deep learning; nonetheless, neural networks, like GPT-3, are very good at language processing.

Prof. Tenenbaum said that it is not a paradox, citing a paper by Martin Schrimpf and colleagues. The paper examined whether neural networks for language processing, particularly GPT-2, can explain activities of a brain region that selectively responds to linguistic inputs. It turned out that neural networks do almost perfectly. This result is consistent with the hypothesis in psycholinguistics that the brain's language system is optimized for predictive processing and has no deep understanding like common sense. In other words, symbolic manipulation and language processing seem to be different.

Both Prof. LeCun and Prof. Tenenbaum, therefore, argued that there is a system bridging the perception, language, and actions. It is the basis of the reasoning and higher cognitive functions that not only humans but also other animals possess.

My Thought

While I might be missing or misunderstanding some critical parts of the discussion, I learned a lot and found the discussion inspiring.

In particular, I like Dr. Silver's suggestion to seek the objective of the mind and his reward-is-enough hypothesis, although my hypothesis is slightly different. In my opinion, the mind's objective is to maximize the reproduction probability of an agent and its descendants, from which (extrinsic and intrinsic) rewards emerge.

Indeed, animal behaviors occasionally deviate from the pure reward maximization principle. For example, human parents sometimes commit a suicidal action to save their children. One might say that it is caused by the mind perceiving the loss of children as a negative reward. Then, why does the mind perceive it as a negative reward? I think it is because the loss decreases the reproduction probability, and the mind obtained the ability to understand it through evolution.

That being said, my hypothesis and Dr. Silver's are almost the same. Mine is that intelligent behaviors arise from the reward maximization principle, but the mind determines rewards for reproduction probability maximization. Conversely, I would say that every creature that tries to reproduce has a mind, regardless of whether it has a central nervous system; many parts of the brain are just tools for the mind to archive its goal.

However, as the number of "tools" increases, it becomes necessary for the mind to have a system dedicated to bridging different brain parts and sending appropriate commands to them, something Prof. LeCun and Prof. Tenenbaum mentioned.

Because of its purpose, the number of such a dedicated system must be no more than one, which might explain why we can pay attention to only one

task at a time. This view is different from that of Prof. LeCun's as there can be multiple world models in this view, but only one manager can listen to them and determine what to do next.

I agree with everyone that the future step of AI and brain science is identifying the neural substrate of the "tool" manager and how it operates.

Discussion on Session 2 talks

Theme : World Model Learning and Inference

Speakers :



Ila Fiete
Massachusetts Institute of
Technology



Karl Friston
University College London



Yukie Nagai
The University of Tokyo
IRCN



Maneesh Sahani
Gatsby Computational
Neuroscience Unit



Tadahiro Taniguchi
Ritsumeikan University

Discussant : Mitsuo Kawato (ATR)

Chair : Hiroaki Gomi (NTT Communication Science Laboratories)

Reporter : Tatsuya Matsushima (The University of Tokyo)

Report

Mitsuo Kawato first provided four points of discussions according to the speakers talk:

- How can achievements of the speakers contribute to neuroscience & AI?
- How do humans achieve learning motor-control from a small sample?
- Finding the essential low-dimensional space for sensorimotor control
- How psychiatric and developmental disorders are related to sensorimotor learning?

In the discussion, the panelists talked about them, which I report in this article.

Learning Generative Models and Model-based Reinforcement Learning

Tadahiro Taniguchi mentioned their achievements on multi-modal concept formulation and language acquisition with generative models described in the talk have been focused on the perception of the cognitive system. However, Taniguchi thinks these models can be meaningful in motor learning as well.

For example, humans seem to generate goals with languages for motor control. In this line of research, Taniguchi argues that the key idea is "control as probabilistic inference", which can be a way of reinterpreting model-based reinforcement learning.

Kawato indicated that it would be extremely difficult to acquire a good internal model from a limited number of samples if there are huge degrees of freedom in the body, although Kawato admits the idea of model-based reinforcement learning itself is notable. Kawato instead pointed out the possibility that the cognitive functionalities help reduce dimensionality drastically and are essential for learning internal models.

Maneesh Sahani mentioned that unsupervised data contribute to learning from a small number of trials and the same underlying schemes can work both for unsupervised learning and reinforcement learning built on the probabilistic inference. Kawato agreed with the idea that reinforcement learning and unsupervised learning take place simultaneously, but added that acquiring a brand new skill requires

finding a new trajectory, which the agent has never experienced, and therefore it is the most difficult part of learning. Sahani suggested that causal structure allows agents to generalize to new environments and trajectories.

Yukie Nagai, who has been working on developmental robotics, mentioned the importance of the limitation of bodies and environments in child learning. Nagai suggested that these limitations can work as good constraints in motor and sensor capabilities. Nagai also pointed out social environments can also be a scaffold for child learning.

Learning Hierarchy

Ila Fiete suggested that learning a new motor action is associated with compositionality and decomposition of representations in the internal cortex and hippocampus. The low dimensional metric variables can be combined to represent a new kind of relationship flexibly. Fiete raised questions about how the brain breaks a problem into the appropriate hierarchies, and how much data is needed to recognize the hierarchical structure.

Karl Friston proposed a view of treating hierarchy as a way of simplifying generative models by minimizing complexity. Friston mentioned that it has a seamless connection with program learning mentioned in the keynote lecture from Josh Tenenbaum, when we see finding a good simple structure having factored representation as a structure learning problem. Friston also added that the framework can be applied to algorithm learning or meta reinforcement learning from the perspective of trying to find the right kinds of structures.

Finding a Good Representation and Neural Structure

Gomi then raised a question on the relationship between finding the essential low dimensional space for sensory-motor control and the neural mechanisms like cerebrum, basal ganglia, and prefrontal cortex. Sahani gave the opinion that distributed representations like successor representations can be well-suited for hierarchical structure because they represent occupancy over the future and linked to

policies. Sahani pointed out that these representations that fit with the inferential view of planning and perception would give us a foundation to work with.

Sahani also mentioned recurrent circuits are necessary to map learning algorithms to neural substrates. The key observation here is that the inferential process around a single variable can be written in a very general form that depends basically on the feedback term in a recurrent network. From the perspective of learning, this scheme can potentially contribute to learning different timescales to find a sensible decomposition of the world.

Sensory-Motor Learning and Psychiatric or Developmental Disorders

Lastly, the discussion goes to relationships between psychiatric or developmental disorders and sensory-motor learning. Nagai introduced their experiments with representational drawing and suggested that hypo- or hyperpriors and difficulties in making hierarchical structures are one of the reasons why persons with these disorders have difficulties in controlling motors and generalizing acquired abilities. Nagai added that modification of certain genes results in an imbalance between excitatory and inhibitory neurons, which is often observed in brains of autism. Nagai supposes that the imbalance also causes hypo- or hyperpriors, which produces too strong or too weak reliance on the predictions.

Friston pointed out that Nagai's idea is consistent with the formulation of the "free-energy principle": if we contrast false inference in these disorders and the view of motor control as "planning as inference", the balance of the prior belief against the sensory evidence can be the reason for wrong inference. These formulations, Friston supposes, can provide explanations on a wide range of disorders like autism, schizophrenia, hallucinations and illusions, and mechanisms for unintentional movement seen in Parkinson's disease. He then stressed that, in this way, insights from machine learning and artificial intelligence can provide useful perspectives on abnormal belief update and psychiatric and developmental disorders.

My Thought

In this session, various concepts, ideas, algorithms, and phenomena were discussed around world-model learning. The main theme of the discussion was “how to learn” the world model, especially, how to learn low dimensional and hierarchical representations essential for sensory-motor control. The unified view for model-learning ranging from bayesian formulation to neural mechanisms and psychiatric disorders introduced by each speaker inspired me a lot.

However, in my opinion, issues on “how to utilize” the learned world model for motor control has also a lot of points to be discussed. Recently, in my research area, the intersection of robot control and machine learning (i.e. robot learning), studies on modeling data in an unsupervised manner using deep generative models become very popular like VAE or GAN, but I feel the utilization of learned models for robot control is less explored. In concrete, one can think of multiple options for using learned models for motor control, for example regarding the model as a simulator that rolls out future prediction of observations or leveraging for trajectory optimization. Therefore, I suppose that discussions and methodologies on model utilization as well as world-model learning are needed for realizing artificial intelligence in the real world. In the discussion session, the topic on “control as inference” introduced by Taniguchi and “active inference” from Friston is potentially related to the model utilization, but not covered so much. I hope I can participate in the discussion about the model utilization at another opportunity.

From the algorithmic perspective of world-model learning, I expect that advances in structure learning would be a key for the next several years. That is because most recent advances on deep generative models are based on engineered graphical models using inductive biases of humans. In order to acquire appropriate graphical models from the data itself, I think we can leverage insights from causal inference or meta-learning. At this point, I agree with the idea of Sahani about mathematical formulation for learning hierarchies and structures behind data streams, and the idea of Friston that minimizing complexity can be

a unified formulation.

Lastly, interpreting the cause of psychiatric or developmental disorders as an imbalance in internal computation for sensory-motor control seemed novel and interested me. I suppose that these insights on mechanisms of the disorders from mathematical formulation may also be useful for discovering a new way of treatment from theoretical aspects. However, since regions of the brain are interconnected with each other, I imagine that careful contrast between clinical evidence and theoretical results of the hypothesis would be required to prevent too much simplification.

In conclusion, I believe this session showed the possibility that the disciplines of different strategies, artificial intelligence and neuroscience, can be contrasted under a unified notion of “world model”. I hope that the content of this session will enhance communications among the panelists, participants, and readers of this article, and lead to a better understanding of intelligence.

Discussion on Session 3 talks

Theme : Metacognition and Metalearning

Speakers :



Matthew Botvinick
DeepMind



Ryota Kanai
ARAYA Inc.



Angela Langdon
Princeton University



Hiroyuki Nakahara
RIKEN Center for Brain
Science



Xiao-Jing Wang
New York University

Discussant : Keiji Tanaka (RIKEN Center for Brain Science)

Chair : Masayuki Matsumoto (University of Tsukuba)

Reporter : Carlos Gutierrez (Okinawa Institute of Science and Technology Graduate University)

After the introduction of the discussion by Keiji Tanaka, Matt Botvinick highlighted the importance of embedding hierarchy for certain mechanisms. In one of his papers, a simple implementation of hierarchy on the architecture led to the emergence of functional hierarchy. That factor can be important in meta-learning.

Xiao-Jing Wang, mentioned 2 important points. The 1st one, while conventional AI implements conventional learning by changing connection weights, an interesting idea is to realize adaptive processes without connection weight updates, where the system still behaving from trial to trial, by connecting neuronal circuits in meaningful weights. The 2nd point is, how to realize gating mechanisms, which still unknown in neurosciences, however, they are starting to have some insights from simple problems by 2 ways: a) trajectories and looking the dynamics, b) changing connection weights in subspaces, then looking the vector field changes. The last one is an interesting angle that would need to be tested experimentally.

About ideas on the interaction of dopamine and the pre-frontal cortex (PFC), Angela Langdon highlighted the question: what constitutes a reward? Perhaps ultimate goals and subgoals? however, looking at the brain, what is considered a reward? This is easy to define in artificial systems but at the brain? Meta-

learning involves generalization and transfer from system to system, thus generalization is not what we think as a “reward”. On the other hand, dopamine is thought as a reward for goal fulfillment. About interactions between PFC and dopamine, PFC is more related to setting the goal, whereas dopamine is related to goal fulfillment, they have different windows of time, this relation seems complex in biological systems.

Ryota Kanai raised a question about what people think of RL and its functions, considering that goal-directedness seems not related to intelligence. How to distinguish between reflecting systems and goal directed systems?

Angela mentioned devaluation in animals, goal changing animal behaviors. The motivational stage in an evolving rodent when they chase water, dictates if they consider that as a goal.

Xiao-Jing mentioned the need of mapping to discrete localized structures. Is it computed locally, or a decision process is taking place within a hierarchy? Ryota mentioned that, about confidence itself, it seems easy to compute. About metacognition or meta-representation, it would be interesting to know what architecture captures meta-representation and how computationally define it.

Matt said, in model base, the decision-making system retrieves the representation, however it still

not a representation of itself. Adding another level of metacognition, such as social cognitions may be important.

Hiroyuki Nakahara raised concerns about the perspective of goal directedness, whether it can be related to model, or simulation, or inference of others behavior? ... it is a broad question. David Silver mentioned that reward may be good enough to generate intelligent behavior. Nakahara continued by considering 2 ways: the only thing to be concerned about is the goal; or a simple reward is the basis for very complex behavior.

Also, he mentioned that meta- "something" seems to include something "without" -meta. The question is, what are we trying to uncover?

Kenji Doya mentioned that, traditionally people assumed modularity and hierarchical structure in the brain, however from Matt and Xiao-Jing work, meta-learning can be achieved without structure. Ryota also presented distributed predictive models, where instead of having metacognition, some coherent space with an overlapping function may be a form of metacognition.

More details were provided by Ryota about his work, where 10 NN are specialized for some tasks, for a given problem. However, he was wondering what area of the brain represents meta-representation of NN or the function of other regions?

Angela mentioned the control over the configuration. In recurrent NN, the configuration is not implicit but embedded in the weights. Metacognition is something like thinking about thinking. Humans can track their thinking, but not necessarily control the configuration of the system.

Xiao-Jing mentioned that he would like to look the learning-to-learn for different kind of problems, involving different blocks of cognition. Having some progress in that front, we can start to question about how we can control.

Matt mentioned that, deep RL will arrive to model-base RL (as shown in his 2018 paper) on explicit knowledge, procedural knowledge so far. But, to achieve reasoning about reasoning, agents may need to include language, or models of themselves (he

introduced the driving drunk example). That still an open question.

Angela added, in biological agents, part of the structure that declares knowledge is social. In human, reward can be the approval from others, and it only exists in the interaction between agents. The social context is part of the meta learning or cognition.

Nakahara mentioned, once we uncover those things, they could be embedded as a meta part or part of the basic systems.

Matt thinks that actually, there is nothing special about meta-learning, it involves any system with large memory, training on many tasks, and with some structure.

Wang mentioned that social cognition is quite interesting. Nakahara said, given the requirement about what we need to learn for survival, (I missed the last part here, just last words before the end of the discussion session).

Below is my personal comments on the topics at this session.

Current AI outperformed the learning of specific tasks using mainly connection weight update and deep layered structures, however, still not able to flexible solve multiple and novel tasks. Meta-learning seems to be a key component on achieving artificial general intelligence. The AI symposium was an unique opportunity to better understand the different, but related, meanings and approaches towards meta-learning and meta-cognition.

One point of view assumes that meta-learning tries to realize adaptive processes in the system (the brain, an AI system) by connecting neuronal circuits in meaningful ways. That adaptation might be realized by gating mechanisms, weight update in subspaces, functional modularity and hierarchies; thought all of those remain poorly understood.

Some comments raised about the need of mapping meta-learning to discrete and localized structures, to understand whether it is computed locally, or a decision process is taking place within a hierarchy. Related to that, mapping what area of the brain represents the function of other regions, also known as meta-representation, could provide a cue on how

“meta” computations are taken place in the brain.

Another view focus on the neuromodulators and their role in the regulation of learning. Particularly, the dopamine is thought as reward for goal fulfillment, however, it is not easy to define in neurosciences what really constitutes a reward. The devaluation of goals is a recent view on this. For example, the motivational stage in evolving rodents when chasing water, dictates if they consider that as a goal. A redefinition of what constitutes goals/rewards and their relationship with neuromodulators on different windows of time seem new research directions. Indeed, the perspective of goal directness still subject of concerns, whether reward may be good enough or not to generate intelligent behavior.

An emergent perspective is the addition of another level of meta-cognition, such as social cognition, for example, specie’s survival as learning or inference of other’s behavior.

Another arising topic considers meta-cognition as something like thinking about thinking. Humans can track their thinking, but not necessarily control the configuration of the system, while in recurrent artificial NN the configuration may be embedded in

the weights. Model base RL showed that some form of representation is retrieved by the decision-making system, however it still not a representation of itself.

In conclusion, traditionally meta-learning assumed modularity and hierarchical structure in the brain, however, the results presented in the symposium showed that it may be achieved without structure. In addition, preliminary results indicate that some form of metacognition can be reached by distributed predictive models in a coherent space with an overlapping function.

The different views on this topic demonstrate its potential and importance on the understanding of distributed learning systems in the brain. The next practical steps, perhaps, should look more to the dynamics, the understanding of trajectories and vector field changes.



シンポジウムの多くの講演と議論は、<https://vimeo.com/showcase/aibs2020> より視聴いただけます。

Discussion on Session 4 talks

Theme : AI for Neuroscience and Neuromorphic Technologies

Speakers :



James J. DiCarlo
Massachusetts Institute of
Technology



Yukiyasu Kamitani
Kyoto University



Rosalyn Moran
King's College London



Terrence Sejnowski
Salk Institute /
University of California San
Diego



Hidehiko Takahashi
Tokyo Medical and Dental
University

Discussant : Kunihiro Fukushima (FLSI)

Chair : Takatoshi Hikida (Osaka University)

Reporter : Yuzhe Li (Okinawa Institute of Science and Technology Graduate University)

I. Interpolating vectors

First, Prof. Kunihiro Fukushima addressed that artificial neural network (ANN) provides design principles for brain-like AI, and is also a powerful tool for analyzing the brain. But the current AI needs big training data to learn, i.e., listen to billions, understand millions, whereas our human being learn quickly, i.e., listen to one, understand ten. Therefore, there appears to be a problem: how to let an AI learn with small data set? Prof. Fukushima proposed a method using “interpolating vectors” with a deep convolutional neural network (deep CNN). Each interpolating vector is determined on a line between two known feature vectors extracted from the training dataset, or by the nearest plane that formed by three reference vectors, or by the hyperplane that made of four feature vectors. This method achieves a data augmentation, not during the training phase, but during the prediction phase. He also proposed a computational trick to reduce the computational cost for calculating the interpolating vectors. This method achieved good learning using a small data set.

Prof. Fukushima's talk inspired Prof. Terrence Sejnowski of the high dimensional problems that was also discussed in previous sessions. He addressed that the interpolating vector method might be an exploring tool for solving high dimensional mathematics problems.

Dr. Rosalyn Moran pointed out that using the interpolating vector as data augmentation reminds her of other data augmentation methods, such as variational auto encoder. These models are able to generate new data, so that can achieve data augmentation.

Prof. Kenji Doya asked if there is any biological interpolation of these interpolation vectors in the brain? And if there is biological evidence shows the internal data augmentation in the brain?

Prof. Fukushima answered that he believes the data augmentation should be represented somewhere in the brain, but he didn't know where it is.

Dr. Sejnowski mentioned there are some psychological experiments showed humans can capture the center of a structure only based on the observations of some random parts, seems they are doing interpolating in the brain too. If we take fMRI while the object doing the experiment, it might give the answer where the interpolating vector is processed in the brain.

Prof. DiCarlo referred to his research in vision system. He found there are mappings between some features in the models and neurons in the brain. For example, the feature extraction is done in the high-level vision system, then can be the basis for the classifier. He thought the interpolating vectors that Prof. Fukushima proposed should be placed in the layer above the feature set, which might be in the

classification layer. There are successful statistical models, but not all of them can be found in the physical groundings in the brain.

II. Evolutionally designed V1 function

Prof. Kenji Doya raised a question to Prof. James DiCarlo: he showed that by replacing the first layer of deep neural network by a V1-like representation, the network's performance becomes much more robust. Then Prof. Doya's question was: is V1 evolutionally designed, rather than trained by sensory experience, for doing the early information coding? Prof. DiCarlo commented whether the mapping comes from evolution, development, or learning is a really challenging question.

Prof. Sejnowski answered Kenji's question: actually there exists evidence to support that V1's function and structure are determined by evolution. There are researches that showed even before the visual system has any experience, there are neurons in V1 that become selective for orientation. Evolution is like learning, but with time step of generation. The visual cortex has more layers than in other cortex and it looks like the function and structure of V1 as the first layer seems already written in the genome. But there are more room of connection plasticity left for higher layers.

Prof. Yukiyasu Kamitani also agreed that the interpolating vector of features should be placed in a higher level. He also mentioned the price equation in population genetics. It explained how phenotype distributes across generations, which follows Bayesian formula. In this way, genome can be treated as memory, and evolution can be treated as some kind of learning. Evolution is not a random process, it has some direction.

III. Reconstruction images from brain activity

Prof. Fukushima asked Prof. Kamitani about his work on reconstruction of images that human subjects are looking at from their fMRI brain activities. Prof. Fukushima wondered if they can also reconstruct moving objects? Prof. Kamitani answered that they are able to decode the direction of moving dots now.

And they are interested to reconstruct movies for future research. Prof. Fukushima was also interested in whether the reconstruction is possible based on the frontal cortex activities. Prof. Kamitani said the decoding from the frontal cortex activities while the subjects are sleeping also reconstructed some images, which might be the scenes the subjects were seeing during their dreams.

Prof. Moren then pointed out an interesting question: it seems that humans generate new data during dreaming, can sleeping achieve a data augmentation? Prof. Kamitani said it might be possible. Maybe there are some generative models in the brain, which might be able to achieve simulation and augmentation of new experience.

Prof. Hidehiko Takahashi added some comments to the the discussion about the high-dimensional space. From a clinical application point of view, he agreed that the clinal data are usually high-dimensional, so the methods for high-dimensional data analysis is an urgently need for their clinical use.

IV. Universality

Prof. Sejnowski raised a discussion on "Universality". He reviewed some research that showed when the recurrent neural networks with different settings were trained in different tasks, even though the trained networks were overall different, the dynamics at some fixed points were similar. He said the universality is the bridge between the ANN and real neuron circuits. If the mathematical model can capture the universal concept, it would be a big breakthrough.

Kenji asked Prof. Sejnowski a question regarding converting ANNs to spiking neural networks. Even though an ANN and the real brain may use different learning algorithms, can the solution they come up be mostly the same? Prof. Sejnowski answered that is their hope. In their explorations they found the more constraints added to the deep neural network, the more architecture and the features of the brain can be captured. He also pointed out that the mathematical models can inspire people design new, more complex experiments.

Prof. DiCarlo added that this is actually what his lab is doing. They are using models to design novel experiments. Sometimes the models are shown wrong, but the experiments may be still partially correct, which leads to designing a new, more accurate model. This circle is the bridge to connect experimental direction and the theoretical direction in neuroscience.

V. Oscillatory dynamics

Prof. Doya asked Prof. Moran about her research using free energy framework to model the dynamics of learning and inference in the brain. Prof. Doya was curious if her model predicts any oscillatory operation in the brain.

Prof Moran said the answer is yes. In the structure of some free energy inference models, it already includes the consideration of gamma wave for forward gain control, and beta or alpha wave for backward control, which may lead to traveling waves. And there are already such applications in analysing the fMRI data and ECoG data.

Prof. Sejnowsk pointed out that there are lots of research explored representations of sensory inputs or motor outputs, but there is a lack of research on how the communications is achieved between different parts in the brain.

Maybe the oscillations is a part of the communications system, which may control the flow of information. We should be thinking about other questions.

Prof. Doya agreed with him, and said many studies assume that, when a subject performs a certain task, the required modules in the brain are activated and connected. But how such global activation and connection can be done biophysically? And what is the computational mechanism behind it? These are still big open problems.

VI. Fixed structure but with flexibility

Prof. Sejnowski added that the brain is not a fixed network. Each brain area is not dedicated only for one particular task. How the brain reconfigures itself for each task in a flexible way is a fascinating question.

But Prof. DiCarlo pointed out that there are a

lot fixed networks in the brain. For example, the retina. Humans may overestimate the flexibility of their brains. The question should be how much flexibility we have in the brain. We can train ANN to do anything we want, even tasks humans cannot do. Most human abilities rely on the evolutionary hardware.

Prof. Sejnowski agreed with the fact that the evolution leaves us a lot of fixed structures. But he also said there is some recent research showing that, even though the retina is well constructed, it is also adaptable, e.g., to illumination aptitudes. So there is a question raised: how could we train a fixed framework (ANN) still have the flexibility?

Prof. Churchland added that the innovations in measuring techniques and analysis changed our view of understanding the brain. Innovation in the measuring techniques revealed that neural actives reflect some changes in the environmental parameters, like Prof. Sejnowski just pointed out. On the other hand, the innovation in analysis methods for population neural actives by using cognitive models gives us a really rich picture of what the brain is doing, which is so different from what we used to focus on.

Discussion on Session 5 talks

Theme : Social Impact and Neuro-AI Ethics

Speakers :



Anne Churchland
University of California,
Los Angeles



Kenji Doya
Okinawa Institute of
Science and Technology



Arisa Ema
The University of Tokyo
Institute for Future Initiatives



Hiroaki Kitano
Okinawa Institute of
Science and Technology



Stuart Russell
University of California,
Berkeley

Chair : Masamichi Sakagami (Tamagawa University)

Reporter : Ho-Ching Chiu (Okinawa Institute of Science and Technology Graduate University)

Masamichi Sakagami started the discussion by introducing some studies on pro-social behaviors: the extent of prosocial-ness and prosel-ness reduced when thinking time increased; also prosel people have higher prefrontal cortex activity while prosocial people have higher amygdala activity. Sakagami suggested that there might be a link between model-free/model-based mechanism with prosel/prosocial-ness, but the detailed neurological mechanism is still unclear.

Sakagami asked Anne Churchland how we can utilize our uninstructed movement activities for brain inspired AI system. Churchland answered that the difficulty is that a majority of electronic signal measured from brain is just low level or movement related activities that we do not care. From an evolutionary perspective, brain is likely developed for movement, and other higher activities just hijacked the machineries and evolved from the movement-oriented brain. Social behavior is also one of basic functions of our brain, since large part of brain activities are devoted to it. Learning how those circuits are repurposed to do other activities will be relevant in developing artificial networks.

Kenji Doya commented that recent advancement of measurement technology might help the dynamic investigation of meaning of different brain activities.

Terry Sejnowski supplemented Churchland's answer that language is one of the examples of

subsequently evolved brain functions, due to the substantial reference of body and space relationships. Sejnowski also mentioned that we are just evolved in the ability to learn to do many things, and practices are needed. Stuart Russell argued, using the example of computer, that despite the first computer was used to decode German, but it turned out the implementation was just logic gates which is vastly general. Therefore, activities other than those biologically oriented do not necessarily have to be based on those evolutionarily designed.

Churchland raised that there are obviously fundamentally different types of motor system like that of ocular system compared to motor and pre-motor cortex. Which suggested there are multiple different motor systems. She also suggested a hypothesis that cognition is at least linked to movement to certain extent, given the difficult of eliminating movements.

Doya asked Kitano whether embodied experience is necessary for one to understand knowledges. Kitano's opinion is that to develop cognition and thought process embodiment might be involved but not essential.

Arisa Ema suggested that it is also important to consider social, institutional and legal kind of influences when we create the brain inspired models.

Russell then stated that the difference between human and machines is whether it has purposes. The

current design of AI systems is based on the notion that intelligence means to have a purpose and to plan how to achieve goals, and so that machines should also have purpose to be intelligent. But this is a mistake because we do not have a way of transferring our entire preference structure to the machine and we end up transferring fragments of it. It might end up we have a machine that is a psychopath, pursuing a distorted version of what we think our objective might be and doing it in a completely single minded way, completely oblivious of anything that is not specified in the objective. What is preferable is some system that is directed entirely towards human's benefit and its extension, like the environment.

Also, it knows that it does not know the objective.

Another philosophical question raised by Russell is how a machine should make decisions on behalf of mankind. How to trade off the preferences of lots of people? One scenario is if someone derive pleasure from the suffering of others, the preference of those persons should be weighted zero since those people have put themselves outside the social contract. Then machine could be designed in a way that it observes humans and figure out their preferences.

Furthermore, Russell asked how many reward is enough? Where does the reward come from? Is it purely biological basis of dopamine which relates to pain, hunger and reproduction? Russell suggested that human may act in a way by observing the others to learn about other's preference and possibly eventually to adopt other's preference. So that the reward system is not just driven by biology but also by culture.

Angela Langdon asked Russell how about the case of addiction, in which there is mismatch of preference between long and short term - in long term the addicts want to integrate to society but short term their preference does not align with their ultimate goal. In this case what should be taken as preference?

Russell respond that there is imperfection in human, so that the mapping of actual behavior to the true underlying preference are very complicated. We need to be able to reverse engineer it for designing AI systems. Also, preferences are not stable and are

modifiable since people are not born with those set of complex preferences. AI systems will take advantage of that, and an example is social media. Whether the notion of preference change is good or bad is nevertheless an open question, since it depends on what the change is.

Ema asked, from a philosophical perspective, what should be the starting question an AI researcher should ask? Sejnowski responded that early philosophy was actually oriented towards physics and science and we are heading in the same direction, but now with the brain. And he mentioned the emergence of a field called neuro-philosophy. Churchland suggested that to discuss this issue it needs to bring together a diverse group of people who is expert in dealing with people whose brains are fundamentally altered, like those with addictions.

Doya commented that there are mechanisms of human society, like democracy and peer reviewing to harness runaway of intelligences. Kitano added that any solutions could not be optimal for all time horizons, like the importance of scientific discovery depends on value systems. An example is that discovery of cure for *C. elegans* cannot get Nobel Prize. Maybe AI should develop a dynamical way for aligning to different value systems.

Russell emphasized that we are not uploading any value system to AI so that the AI should be constantly uncertain about what people want and how to trade off those preferences.



田中 啓治

理化学研究所 脳科学総合研究センター
シニアチームリーダー



疋田 貴俊

大阪大学蛋白質研究所 高次脳機能学研究室 教授



田中 疋田先生どうぞ、今日是对談よろしくお願ひいたします。私がホスト役ということなので、主に聞き手に回りたくて思っております。疋田先生から面白い話を、若い人に参考になるお話を、引き出せればと思っております。

疋田 よろしくお願ひいたします。

田中 対談は計画班員の中で順にやってる訳ですけども、計画班員の中での疋田先生の特徴は、精神科で研修をなさって、精神科がバックグラウンドにあることじゃないかと思ひます。精神科に入られた動機とか精神科の経験が今の研究にどう役だった、どう影響してるかとか、そういうお話から少し聞かせていただけませんか。

疋田 ありがとうございます。私は京都大学の学部生の時から基礎医学研究に進むことを考えていましたが、精神科には脳の疾患として最も興味を持ち研修を行いました。京都大学医学部附属病院の精神科で研修医をした後に、京都大学大学院の基礎医学教室に進学をして、その後は神経科学に従事しています。しかし現在も、土曜日だけですが、精神科医として臨床をしています。精神科臨床は、直接研究の対象とはしていませんが、精神疾患を対象に含む私の研究テーマに多くの影響を与えています。精神疾患はまだ病態が未解明で、研究対象としても魅力が大きい分野です。

田中 神経内科の病気は、パーキンソン病やアルツハイマー病など、ある程度疾患の原理が分かって、治療に向かっているんですけども、精神疾患の方は色々な仮説は出てきているけども、それが治療に結びつくという見通しはまだ全然ないと言っているのではないかと。

疋田 その通りだと思います。

田中 土曜日に精神科で診てらっしゃるっていうのは、大学病院で診てらっしゃる?

疋田 精神科クリニックの外来で、何人かの患者さんを診ています。

田中 クリニックではうつ病の患者さんが多いんですか。

疋田 はい、うつ病の患者さんを多く診ています。

田中 激しい統合失調症の方とかは。

疋田 統合失調症の患者さんは関連病院の入院施設で、時々診ています。それはやっぱり大学病院の方がより経験がありました。

田中 重篤な患者さんへの治療は主に対症療法だというふうに聞いていますが。

疋田 統合失調症の病態が未解明ですので、症状に応じた治療となっています。しかし、私が研修した時と比較して、治療薬の種類も増えてきて、環境は良くなってきています。早期に治療を開始できるとほぼ普通の社会生活が可能になってきています。基礎研究から統合失調症の病態解明を進めることで、今後治療法の開発にも貢献していきたいと思ひます。

田中 症状を抑えることによってそれ以上の疾患の進展を抑えることができるようになります。

疋田 薬がない状態ですと病状が進行してしまうことが多いですが、薬をきっちり服用していると進行を止めて、その段階での生活を継続することができます。

田中 それから、先生のご経歴を見ると、大阪バイオサイエンス研究所で研究員をなさった後に、京都大学で武田薬品工業との産学連携の講座でPIを始められたというように拝見しますけども。

疋田 はい、京都大学大学院医学研究科メディカルイノベーションセンターで行われました武田薬品工業と京都大学との協働でのプロジェクトの中で、統合失調症の創薬を目指すプロジェクトのPIの一人として、特定准教授を5年間務めさせていただきました。創薬となりますと長い道のりになりますので、その最初の段階の5年間を取り組んだ形になりました。

田中 疋田先生は学生時代から大脳基底核の研究をし、大

阪バイオサイエンス研究所の研究員の時も大脳基底核の研究をされたと思うのですが、武田薬品工業との連携でも大脳基底核中心に研究されたのでしょうか？

疋田 はい、神経回路の研究から、精神疾患の病態をターゲットにすることを考えておりました。大阪バイオサイエンス研究所の前に、ジョンズホプキンス大学に留学してまして、澤明教授の研究室で、精神疾患のモデルマウスの作製と解析に従事していましたので、精神疾患モデルマウスと大阪バイオサイエンス研究所での大脳基底核の神経回路の研究を組み合わせると、そうするとモデル動物の神経回路を調べることで、何か新しい創薬の目が見つからないという研究を続けていました。

田中 ジョンズホプキンスの時には有名なDISC1のモデルマウスを作られた。そのマウスと大脳基底核の関係を研究されたのですか？

疋田 はい、精神疾患モデルマウスと大脳基底核神経回路を柱に研究してきました。精神疾患の創薬での最大の障壁は薬効を確認するためのバイオマーカーの欠如です。モデルマウスは幻聴や妄想などの精神症状を発症する訳ではありませんので、何を指標にして薬効をみるかが確立していません。そこでモデルマウスの認知機能に焦点を当てて、精神疾患の創薬のためのバイオマーカーとなりうるかを調べています。また、認知機能の回路機構の解明も大事ですので、大脳基底核を中心に研究しています。

田中 そうですか。例えば統合失調症であっても陽性症状や陰性症状が出る前に、もっとその前から認知障害が出ている。それをモデルマウスで実現することによって創薬のターゲットにしようと、指標にしようというアイデアですね。具体的にはどんな認知障害ですか？

疋田 ヒトにもトランスレーション可能なマウスの認知課題を設定して、成績や反応速度などの指標でモデルマウスの認知障害を調べています。タッチスクリーンに図形課題を提示し、モデルマウスではどの指標でどのような変化を起こしているのか、さらにはそれぞれの指標にどのような神経回路基盤があるのかを調べています。

田中 統合失調症の回路基盤には何か手がかりがあるのでしょうか？

疋田 最近の研究では、統合失調症も、遺伝背景の強い自閉症などの幼年期の精神疾患と同様に神経発達障害として捉えられてきています。統合失調症では、まず遺伝的な影響による神経回路の発達の障害があり、さらに青年期での社会的ストレスが加わることによる遺伝環境相互作用による発症が想定されています。私はこれまで大脳基底核の神経回路機構を研究してきて、大脳基底核の直接路と間接路の2つの回路が違う役割を担っていることを見つけてきました。つまり通

常ではうまく2つの回路を使い分けて、その場面にふさわしい認知行動をとっていると考えています。そうしますと精神疾患では、ドーパミンの機能異常の結果として、神経回路自体が変化しているのではという仮説で色々調べています。

田中 そうですか。統合失調症というとなんとなく前頭葉、前頭前野に関わるような障害じゃないかというふうに考えられていますけど、大脳基底核が深くかかわっていることでしょうか。

疋田 前頭前野は大脳基底核とループ回路を構成していますので、前頭前野と大脳基底核の一方というよりは、ループ回路で説明がつけられるのじゃないかと考えています。

田中 大脳基底核というと、前頭葉ループ以外に運動野ループもあるわけですが、統合失調症で運動野ループに異常は起こらないのですか。

疋田 運動野ループの異常はあまり言われていないですが、薬の影響は運動野ループを含めた回路レベルで見ると思っています。

田中 そうすると薬も大脳基底核関係の薬が多いのですか？

疋田 基本的にはドーパミンD2受容体拮抗作用を持つ薬が基本で、ドーパミン放出の制御を行っていると考えられています。しかし、大脳基底核ループ回路でも作用するのではと考えています。

田中 ドーパミンは大脳基底核に一番強く投射しますが、マカクで見ると前頭前野にも強い投射があるので、そのドーパミンがその直接前頭前野に投射する経路の影響はどうなんでしょうか？

疋田 はい、ドーパミン神経は前頭前野にも投射しています。計画班員の松本正幸先生の研究でもドーパミン神経の活動様式は多様性を持っており、投射先によってもドーパミンの役割が異なることがわかってきています。前頭前野に直接投射するドーパミンも精神疾患で重要な役割を持っていると考えています。

田中 ドーパミン細胞のうちの大脳基底核に主に投射する細胞と、それから前頭前野へ直接投射する細胞は別の細胞なのですか、それとも同じ細胞なのですか？

疋田 他の研究室での研究からも、大脳基底核に投射する細胞と前頭前野へ投射する細胞は同じ神経核からのドーパミン細胞であっても、それぞれ別の細胞ではないかと考えられています。

田中 こうしてお話を聞いていると、やっぱりドーパミン系というのは精神疾患を攻めていくときも一つのやっぱり非常に重要な要素である？

疋田 はい。精神疾患とドーパミン系の研究は古くからなされてるのですが、まだまだ新しいことが見付かっていく領域だなと感じています。

田中 そうですね。じゃちょっと話を少し転じまして、疋田先生というと、精神科出身ということと、もう一つは、中西重忠先生の研究室でPhDをお取りになって、分子生物学のバックグラウンドも非常に強くなる。この研究班の中では比較的少ない存在だと思うんです。私が学生だったころは分子生物学が始まってから、まだ歴史の浅い時期で、私は学生の時に大澤文夫先生、中西先生よりもさらにもっと年上の、分子生物学のパイオニアの一人の先生の授業を受けて大変感激しました。大澤先生がいつもおっしゃってたのは「機能あるところにタンパク質あり」。何か特別な機能があったら特別なタンパク質があると彼は説いたのです。脳でそんなことがあるかって私は反発したんですが、でも先生の授業は凄く面白かった。現在の脳科学でも、分子機構や疾患の分子病態を研究するという立場と、もう一つ分子遺伝学を道具として使う、神経回路の機能を解きほぐしていくための道具として使うという2つの立場があるように思うのです。そこら辺は、今、疋田先生は、どういうイメージを持っていらっしゃいますか？

疋田 はい。最近の神経科学は分子をツールとして使って神経回路機構を解き明かす方向性が主かと思います。実際、我々もそこは一生懸命やっています。一方で、私が創薬プロジェクトに参加していた経緯から分子病態を明らかにし、創薬の分子標的を探る方向性も重要視しています。

田中 これから脳科学の発展につれて、そのどっちが中心になってくるとお考えになりますか？

疋田 分子ツールを使って調べることは神経科学としてまだ多く残されています。一方で、回路ごとあるいは、一細胞レベルの分子機構を調べる技術も発展してきていますので、そのような方向性が出てくるのではないかと考えています。

田中 そうですね。私は分子生物学のバックグラウンドは弱いですが、それでも最近のレビューを読むと、神経細胞も分子的に非常に多様であると。

疋田 そうですね、はい。

田中 大脳皮質の領野ごとにも遺伝子発現が随分違う。そういうことを聞くと、やっぱり大澤先生の言われたことも、まだまだ見えてないけども、これから見えてくる可能性もあると？

疋田 はい、そのような分子機構の違いがどのように神経回路を制御しているのかを見ていきたいと思っています。

田中 ところで、そういう分子生物学の考え方、機能の裏には独特のタンパク質ありという考え方と、我々の領域を特徴づけるネットワークの考え方や人工知能の考え方の関係。同じ素子であっても繋がり方によって機能が違ってくるとというのがネットワークの考え方だと思うので、ふたつは、ある意味では、両極端にあると思うのですが、この計算論的神経科学や人工知能と先生のご研究の関係についてはいかがですか、どんな

ことを考えていらっしゃいますか？

疋田 私が分子生物学出身の脳科学者なので、どうしても物質中心のところから研究をスタートしているのが現状で、人工知能とか計算論的精神医学を今学んでいる状況です。本新学術領域に参加させていただいて、もうすぐ新学術領域の研究期間は終了を迎えるところですが、領域内外の共同研究により人工知能や計算論的精神医学への取り組みをようやく始めることができている。動物の行動などの神経回路をしっかりと見たと、人工知能に取り組みたいと思います。私の主な研究対象でありますドーパミン系は、強化学習と関連が深いので、脳科学の側面から人工知能分野にも貢献できたらと思っています。

田中 分子の振る舞いも、強化学習などの理論を理解してないと理解できないですよね？

疋田 そうですね。分子の振る舞いにも非常に興味があります。私は蛋白質研究所に所属していますので、神経科学だけではなく蛋白質科学の領域でも貢献したいと思っています。例えば病気の原因となるような精神疾患の分子変異が、どのようにタンパク質レベルで振る舞いの異常を引き起こすのか、次にこれが神経回路レベルでどう変わっているのかとも見ていきたいと思っていますが、これからになります。

田中 神経細胞ごと、領野ごとにも凄く違う発現をしているタンパク質と、人工知能に対応するようなシステムの機能がどう関係してるかというのは、これから10年20年の話題だと思うんです。これからの10年20年の抱負はいかがですか。

疋田 2017年に大阪大学蛋白質研究所で高次脳機能学研究室を開始しました。定年まで20年ありましたので、20年かかるテーマを進めていけたらと考えています。脳科学はそのようなテーマに相応しく、本新学術領域に参加して研究テーマをさらに広く展開できていることを非常にうれしく思っています。

田中 そろそろ時間になりました。先生のこれからの研究のご発展を見守り続けます。

疋田 ありがとうございます。今後ともご指導をお願いいたします。

音声命令を用いたナビゲーションのための場所概念に基づく確率推論

谷口 彰（立命館大学 情報理工学部 特任助教）

論文タイトル: Spatial Concept-Based Navigation with Human Speech Instructions via Probabilistic Inference on Bayesian Generative Model

著者名: Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi, Tetsunari Inamura

掲載誌情報: Advanced Robotics, Vol. 34, No.19, pp.1213-1228, Sep, 2020.

DOI: 10.1080/01691864.2020.1817777

はじめに

人間の生活環境で動作するロボットは、適応的に場所に関する知識を学習・活用することが求められます。これまで我々の研究では、図1（上側）のように場所概念と呼ばれる場所に関するカテゴリ知識を、ロボットの知覚センサ情報と人の音声指示から獲得する確率的生成モデル SpCoSLAM を提案してきました。今回の論文では、SpCoSLAM のモデルにおいて Control as Inference (CaI) と呼ばれる推論の枠組みを導入し、音声命令が与えられた際の目的地までの軌道計画を行う手法 SpCoNavi を提案しています。具体的なタスクとしては、図1（下側）のように「ミーティングスペースに行って」などの場所の移動に関する命令を受け、ロボットは現在位置から目標状態への行動決定を行います。この手法の主な特徴として、(i) ロボットが自らの経験をもとに自律的に獲得した場所の概念・語彙を活用すること、(ii) ゴールとして地図上の座標点を指定する必要がなく、人の音声命令から目的地までのナビゲーションが可能であることが挙げられます。

CaIに基づくパスプランニングの定式化

CaI では、強化学習や最適制御を始めとする行動決定問題を確率モデル上での推論として捉えることで、統一的な確率推論の枠組みで議論できることが利点です。行動と状態の遷移を表す Markov Decision Process (MDP) のモデルに最適性変数 o_t を導入することによって CaI は実現されています。最適性変数 o_t の導入により、軌道の事後分布 $p(\tau|o_{1:T})$ における最大事後確率の推論がプランニング問題に帰着されることが知られています。

提案手法では、人からの音声命令 y_t が与えられたときの動作 u_t と位置 x_t の軌道 $\tau = \{u_{1:T}, x_{1:T}\}$ を表す事後確率が最大となる動作系列 $u_{1:T}$ を推定する問題として定式化します。目的

関数となる確率分布が最大値をとるときの動作系列を求める式は以下で表されます。

$$u_{1:T} = \underset{u_{1:T}}{\operatorname{argmax}} p(\tau|y_{1:T}, x_0, \Theta_G)$$

このとき、 T は計画区間、 x_0 はロボットの初期位置、 Θ_G は SpCoSLAM によって学習済みの場所概念のパラメータを表します。具体的には、地図のセルごとに動的計画法 (Viterbi algorithm) によって確率値を計算し、軌道の推定を行います。

SpCoNavi では、音声命令 y_t をゴール信号を表す観測変数であるとみなしており、これは CaI の視点で見ると最適性変数 o_t に対応しています。つまり、与えられた単語がより観測されるであろう場所に向かうことが最適であることを意味します。CaI に基づき強化学習との対応を考えると、(報酬関数を手動で設計することなく) 音声命令における目的地を表す単語の条件付けに依存して報酬関数が自動的に切り替わっていることがわかります。

場所と単語の多対多対応に関する効果的な応用例

本論文の実験の一つとして、場所と名前の多対多の対応関係を考慮に入れたプランニングの実証を行いました。図2(a) - (d) は、それぞれ異なるシチュエーションを示しています。図2(a) はいくつかの候補地から典型的な（主によく使われている）場所への移動、(b) はある場所に対して複数の名前が存在する場合、(c) は音声命令の中に2つの単語が含まれる場合（AND 条件）、(d) は音声命令の中に2つの単語が含まれている場合（OR 条件）になっています。カラーマップを見ると、与えられた単語に応じて対数尤度が変化している様子がわかります。結果として、4種類のどのケースにおいても適切にパスプランニングできることが示されました。

おわりに

本研究では、ロボットが自律的に獲得した場所概念を用いた音声命令からのパスプランニング手法を提案しました。提案手法は、従来の多くのパスプランニング手法とは異なり、人との音声によるインタラクションから目的地までのナビゲーションが可能となります。今回紹介した実験では、提案手法が実行できる様々なシチュエーションについて示しました。この他に手法の詳細や実験結果など、今回紹介していない部分に関しては元論文をご参照ください。

今後の展開として、トポロジカルマップ（グラフ構造を持った地図）への応用や、リアルタイムナビゲーションのための計算量の削減、能動推論（Active Inference）との融合などが挙げられます。また、このような場所概念に関する確率的生成モデルや確率推論と、空間認知や場所の認識に関わる脳部位である海馬体との関係性や対応付けについても議論を進めていく予定です。

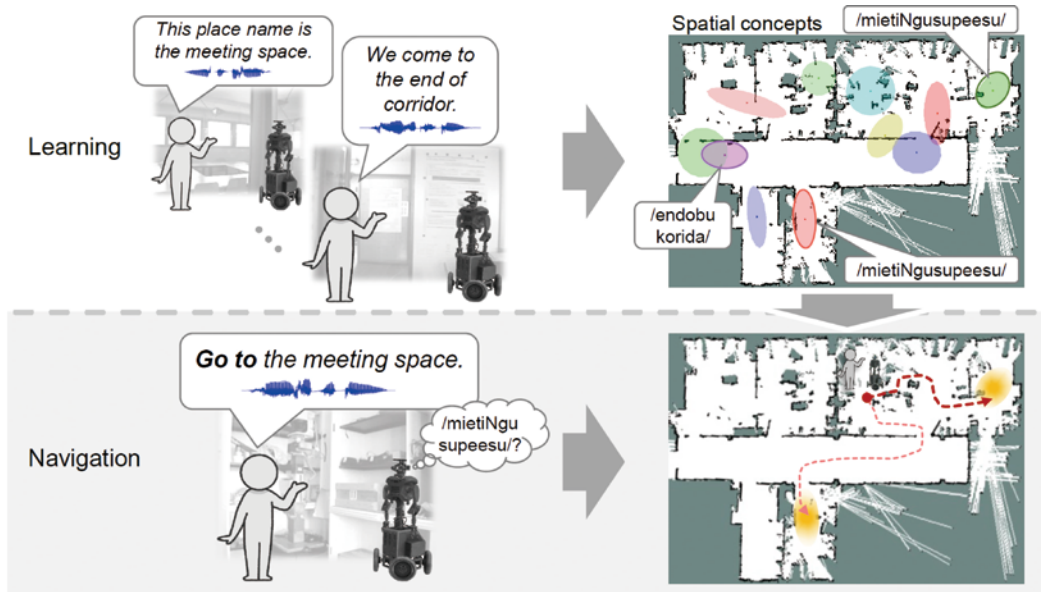


図1：タスクの概要図。(上側) 場所概念の学習フェーズ、(下側) ナビゲーションフェーズ。提案手法は、学習された場所概念と音声命令から目的地までの軌道を探索する。

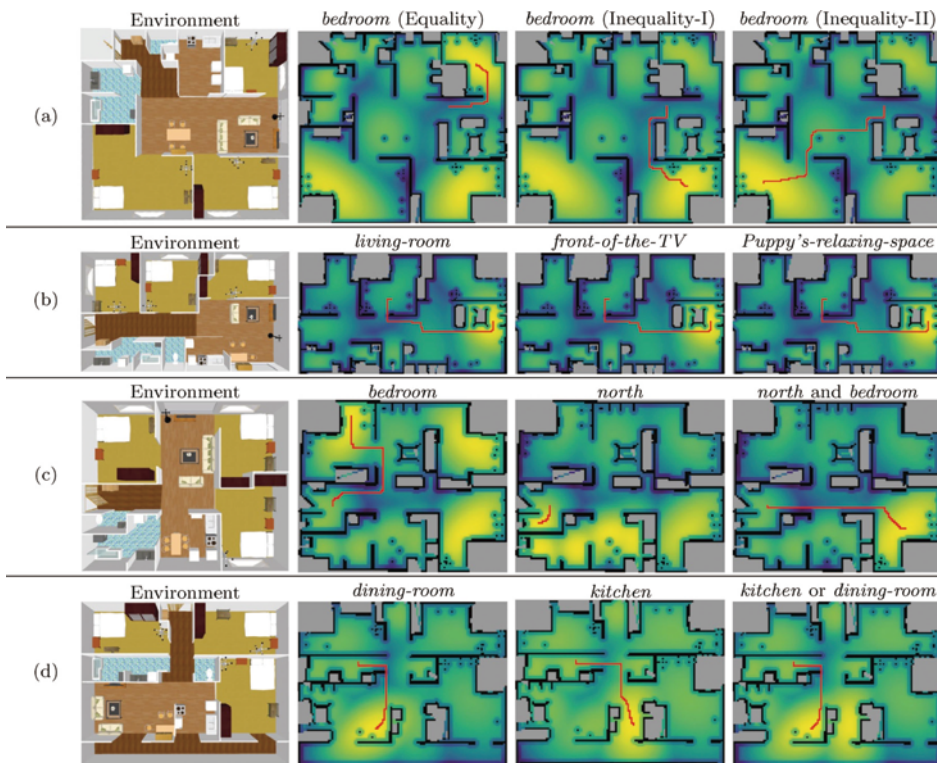


図2：様々なシチュエーションの音声命令によるパスプランニングの結果。赤線が推定された軌道。カラーマップは対数尤度の値を表す（黄色が高く、青色が低い）。

IROS2020 参加記

長谷川 翔一（立命館大学 情報理工学部 谷口研究室 博士課程前期課程 1年）



2020年10月25日から12月24日にかけて、オンデマンド形式で行われたIROS2020(IEEE/RSJ International Conference on Intelligent Robots and Systems 2020)を聴講してきました。IROSはIEEEと日本ロボット学会(RSJ)が開催するロボットに関する国際学会であり、他のロボット系学会のICRAやRSSなどと並ぶ世界的に有名な学会です。2020年度は新型コロナウイルス(COVID-19)感染拡大のため、WEB上での一般公開される形となりました。私は今回、自身の研究分野における最先端の研究調査のため、本会議の参加を行いました。

IROS2020は当初、アメリカ合衆国のラスベガスで開催される予定でした。しかしながら、世界的に感染が拡大したCOVID-19の影響懸念のため、WEB上でのオンライン開催に急遽変更されました。また、IROSへの参加登録費が無料となり、参加登録を行えば、誰でも発表者の講演の様子や論文を閲覧することが可能になりました。発表者の講演が事前に録画されているため、聞き逃した部分は何度でもビデオ再生することができ、オンライン開催ならではのメリットが感じられました。

IROS2020はTechnical Talks、Workshops and Tutorials、Robot Competitionの大きく3つのパートに分かれています。Technical Talksでは、主に基調講演や論文が採択された人の発表を閲覧することができます。またTechnical Talksは12のカテゴリに区分され、各カテゴリの中でさらに複数のセッションが含まれています。自身の専門領域外の研究も多くあり、興味のある研究発表を気軽に閲覧することができました。次にWorkshops and Tutorialsでは、発表者だけでなく聴講をしている参加者も一緒に議論できる場

になっています。Zoomなどのビデオ会議ツールを用いて、別日程で議論の場を設けていたものもありました。最後にRobot Competitionでは、実際にロボットを動作させ、物体を把持するタスクなどを行い、技術力の高さを競い合っていました。また、Competition参加を通じて、チーム同士の技術に関する議論がZoomなどで行われていました。

印象に残ったものとして、Plenariesで講演をされていた東京大学の長井志江先生が挙げられます。人の認知発達を知る手掛かりとして、Fristonらが提案した予測符号化をはじめに説明していました。これは、環境から受け取った感覚信号と脳の前測器が予測した信号間の誤差を最小化し、脳の前測器を学習していく理論です。この理論に基づいて、人の社会性の発達と自閉スペクトラム症(ASD)を例とした認知発達の多様性に焦点を当てていました。人の社会性の発達では、14カ月の乳幼児が人を助ける行動を例として挙げ、ミラーニューロンが、行動の動機を支えているのではないかと提起していました。研究では、ロボットに感覚情報から物体に手を伸ばす行動を学習させ、学習したネットワークから未観測の感覚情報予測を可能にしており、ロボットに感情のカテゴリ獲得を可能にする研究も行っていました。次に認知発達の多様性では、社会性障害を持つASDを予測符号化によって発生原理を解明しようとしていました。上記のことを実現するために、ベイジアンベースのRNNにより予測符号化を構築し、予測精度に関わるパラメータを変動させることで、ASDの性質が再現されるかを描画表現タスクで検証していました。結果として、パラメータを変動させると、変化させない場合と比較して、描画表現に明確な差が見られました。

最後に、オンライン開催ではありますが、初めての国際会議の聴講を通じて、自身の研究分野に関する情報を得る良い機会だったと思います。他にも興味のある論文が複数あるため、引き続き論文を読み進めたいと考えています。また、IROSに限らず自身の研究を国際学会へ投稿し、採択されるために研究を取り組んでいきたいと思っています。

CoRL2020 参加記 Conference on Robot Learning 2020

山森 聡（京都大学ATR連携講座 博士後期課程2年
ATR脳情報通信総合研究所 ブレインロボットインタフェース研究室）



©2020 Conference on
Robot Learning

CoRLはロボット工学と機械学習の交差点を焦点とした国際会議であり、今年で4年目を迎える比較的新しい会議です。両分野における近年の盛り上がりを反映し投稿数も年々増加傾向にあり、独自性のある注目されるコミュニティです。今年にはマサチューセッツで開催される予定でしたが、残念ながら新型コロナウイルス感染症の影響から現地での開催は行われず、オンライン会議での開催となりました。

オンライン会議の形式はPheedLoopを利用したポータルサイトによって提供されており、個別のセッションごとにウェブ会議サービスであるZoomへアクセスするという形式になっていました。このPheedLoopのポータルサイトがなかなか便利で会議スケジュールの通知機能や他の参加者とのコミュニケーションツールなども備わっていました。また本会議自体は単一のZoomリンクを通して基調講演やチュートリアル、オーラル発表などが提供されています。これらの講演は音声のみで講演者のみに許可されるため参加者の質問はチャット上で司会者へ送信後に司会者が代理で質問するという形式であり直接参加者が質問できない状況でした。しかし事前に質問を整理できるためか、むしろ質問の数は多く活発な議論がなされている印象を受けました。一方で例年のポスターセッションは各発表者に割当てられたZoomを通して発表者と参加者間での直接会話形式で提供されておりこちらでも活発な議論がなされていました。すべての発表に関して事前に3分間のスポットライトが動画共有サイトYouTubeに投稿されており、ポスターセッションというよりは事前に閲覧したスポットライトや論文における疑問点を直接質問する場として活用されていました。

実際の会議における研究動向ですが今年の投稿数は475本で採択率は34.74%でした。投稿数は昨年から20%上昇しており分野および学会への注目度が伺えます。また、採択率は低下傾向にありましたが30%台を維持する傾向が見て取れます。論文のキーワードとしては投稿論文および採択論文ともに20%前後の割合で強化学習が含まれており引き続きロボット工学における強化学習への強い関心があることが伺えます。採択率が最も高い研究キーワードは80%の割合で“Legged Robot”であり、これは効果的な学習手法が検証対象に脚ロボットを選択していることを示唆しています。同様にマニピュレーションや模倣学習を応用的に扱う論文も増加しており、実問題を志向した研究が評価されています。

ここからは個別の発表について取り上げます。

始めに“SMARTS: An Open-Source Scalable Multi-Agent RL Training School for Autonomous Driving”

は自動運転におけるマルチエージェント強化学習（MARL）の検証環境を提供する上海交通大学とHuaweiの研究です。自動運転は機械学習の適用が期待される分野ですが、これまで複数のAIドライバーによって操作される車が環境と相互作用しながら学習する環境が未発達でした。SMARTSによって今後新たなMARLの手法が提案された際に統一的で軽量な検証環境を提供されました。本研究はBest System Paperを受賞しておりベンチマーク系の研究もCoRLの一つのトピックです。

続いて、“Learning Latent Representations to Influence Multi-Agent Interaction”はスタンフォード大学から提案された協調制御を実現する強化学習の手法です。人とロボット間の協調制御において人はロボットとは独自に動くためロボットから見た場合に非定常な環境を持つマルチエージェントシステムとなり単純な強化学習的枠組みでは扱えません。本研究では相手（人間）の行動を潜在的な方策と潜在空間を組み合わせた戦略レベルの方策の2つに階層化し、ロボットが相手の潜在空間表現を学習することで、相手の戦略を予測し適応した行動を選択できるように学習する手法です。これによって相手の行動が完全には予想し得ない場合でも戦略レベルで協調し相手の行動を変化させることができます。

最後に“Accelerating Reinforcement Learning with Learned Skill Priors”は南カルフォルニア大学から提案された階層的強化学習による別タスクでの学習速度の高速化に関する研究です。強化学習ではデータサンプルに対する学習効率が課題ですが、この研究では以前に学習した方策における潜在的スキル空間を学習し、新たなタスクではスキルを選ぶ方策を学習することで学習効率を高速化します。ここで、スキルの組み合わせを以前選ばれた組み合わせに近いものを選択することで効率的な探索が可能になります。

会議全体では同様の深層学習におけるオートエンコーダーを用いた潜在空間学習を方策だけではなく、報酬や状態空間に適用する例が多く見受けられました。深層学習のロボット工学への応用が多く見られますが、制御理論を基にした安定性の保証やリーマン幾何と動力学モデルを駆使したプランニングなどの発表も注目されています。今後はこれらの分野統合や体系化が志向されることを予感させるものでした。

今回の参加は本新学術領域による支援を受けています。貴重な機会を頂けたこと、感謝致します。

JNNS2020 Conference Highlights

by Carlos Enrique Gutierrez, post-doctoral scholar, Neural Computation Unit, OIST



From December 2nd to 5th 2020, the Japanese Neural Network Society conducted its 30th annual conference, the JNNS2020, originally planned to take place in Sendai city, in an online format (<http://www.nanospin.riec.tohoku.ac.jp/JNNS2020/>). Posters and oral talks were accessible on-demand, by videos and chat panels, including live oral sessions as well.

Keynotes focused mainly on motor and sensory systems of the brain, disclosing interesting results. Evidence of the hierarchical organization in the medial and lateral prefrontal cortex and its implication on action-control captured interest (Hajime Mushiake); as well as the introduction of brain's hyper-adaptability, interpreted as its adaptive capability under impairment of sensory, motor and cognitive functions, with emphasis on the population aging issue (Jun Ota).

Furthermore, intriguing experimental results, from humans and monkeys, of the default mode network role were featured, together with the identification of the precuneus as the key area representing the "here" (the visual world in front) and the "now" (the present time, determined experimentally by verbal stimulation). Discussions on that topic examined links with consciousness (Shigeru Kitazawa). Finally, a particular keynote talk discussed the evolution of science and technology and its impact on society, by a chronological timeline starting from the 17th century to present day. Special highlights on the strengths of Japanese technology as well as present and future

challenges were debated (Osamu Sakura).

Engaging brain-hardware related talks displayed not only the usual advantages on reduction of power consumption and improvement of computing performance, but advances on hardware architecture. For instance, hardware inspired by the hippocampus with strong implications on memory functions, such as memory storing and recalling (Takashi Morie, Katsumi Tateno, Hiroki Nakagawa and Kensuke Takada). Besides that, noise, spike generation and cell function mimicked by nanomaterial-based molecular devices captured special attention, since its potential as "new land" for the development of neuromorphic devices (Megumi Akai-Kasaya and Tetsuya Asai).

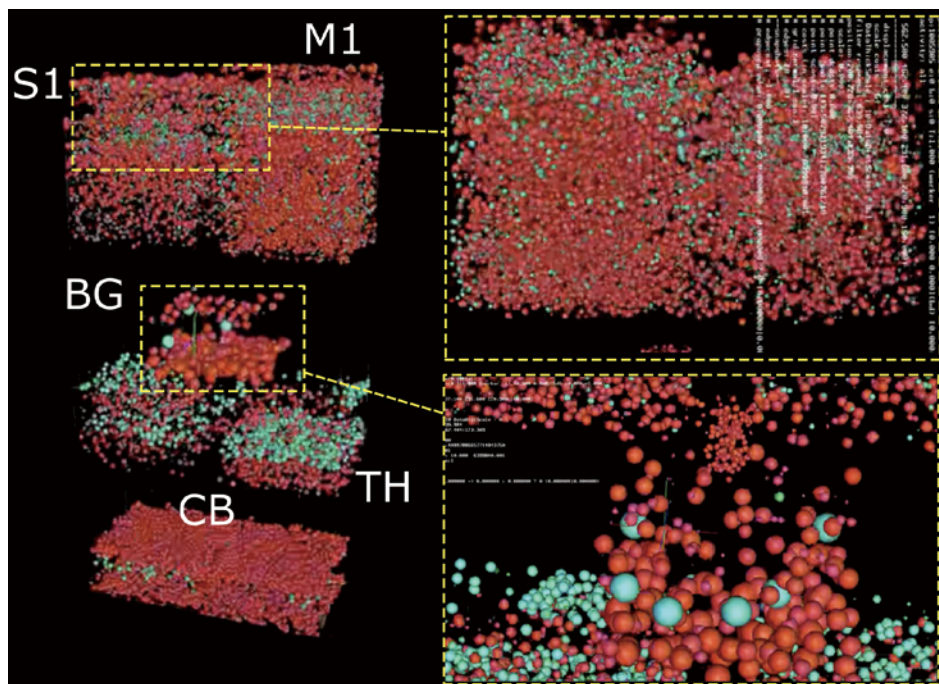
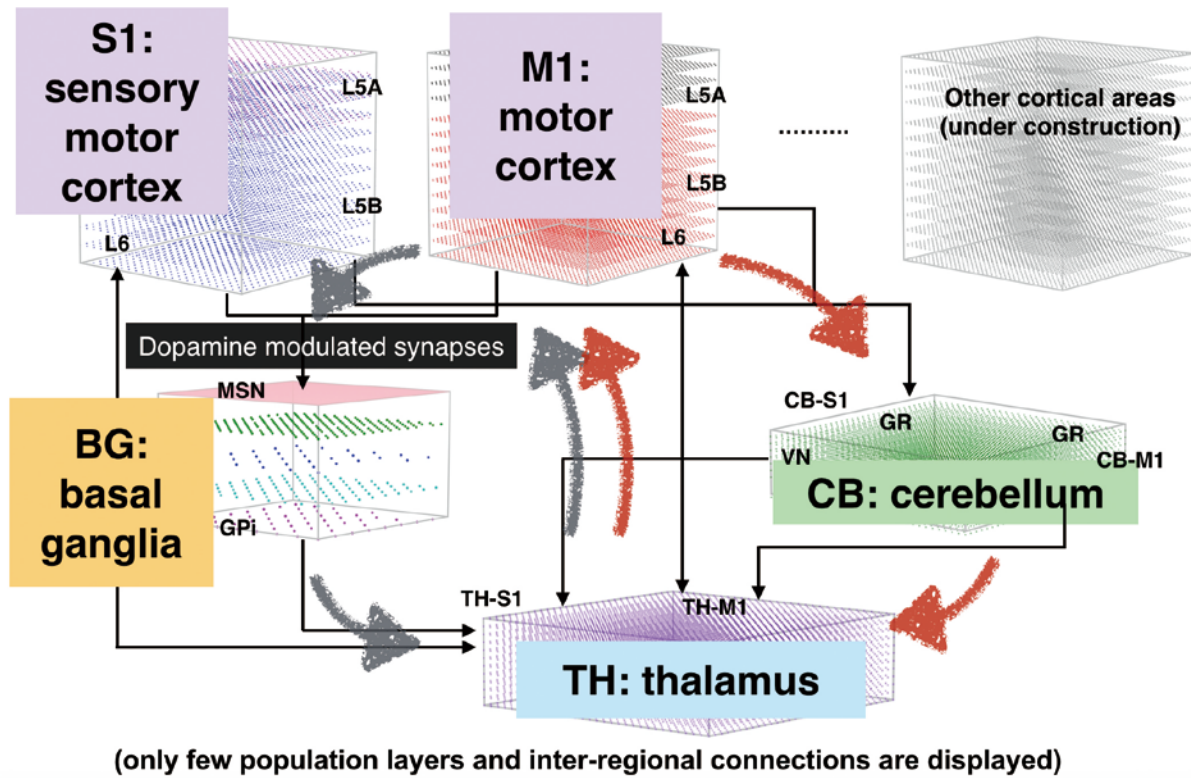
Further sessions presented a wide variety of topics, from which I would like to highlight findings on spatial features related to motor control. Such is the case of arm motor skills dependency on the coordination between eyes and arms, with links to motor regions during planning and the representation of motor memories (Naotoshi Abekawa, Sho Ito and Hiroaki Gomi); as well as conclusions on the standing posture contribution on leg control during walking, performed by comparison of muscle activations for both modalities (Chika Yamane, Tomoko Hioki, Shoko Kaichida and Jun Nishii). Decision making models were introduced likewise, for example, a reward and curiosity decision-making model for a 2-choices task based on the popular free-energy principle, analyzed conservative-explorative behaviors based on curiosity level (Yuki Konaka and Naoki Honda).

Our group had the opportunity to present a brain simulation of the resting-state neural activity in a loop circuit of the cerebral cortex, basal ganglia, cerebellum, and thalamus, using the state-of-the-art "NEST" neural simulator. This work is a collaboration across several labs specialized on brain modeling, and targets the investigation and understanding of the dynamic nature of the whole-brain network. The model's code and simulation were

opened to the scientific community (<https://github.com/oist/whole-brain-jp>), since its aims to serve as a platform for further investigation of brain functions. Moreover, large-scale simulations on the former K-computer were introduced, together with full-brain

simulation challenges on Fugaku supercomputer.

On the last day of the conference, a satellite symposium, the 9th RIEC International Symposium on Brain Functions and Brain Computer, was held with presentations in English.



Top: Graphical outline of the integrated whole-brain model. Its network structure makes up more than 90% of neurons in the mammalian brain. Bottom: Screen capture during a large-scale simulation displaying the neural activity of several types of spiking neurons.

NeurIPS2020 参加記

古田 拓毅 (東京大学 松尾研究室 修士1年)

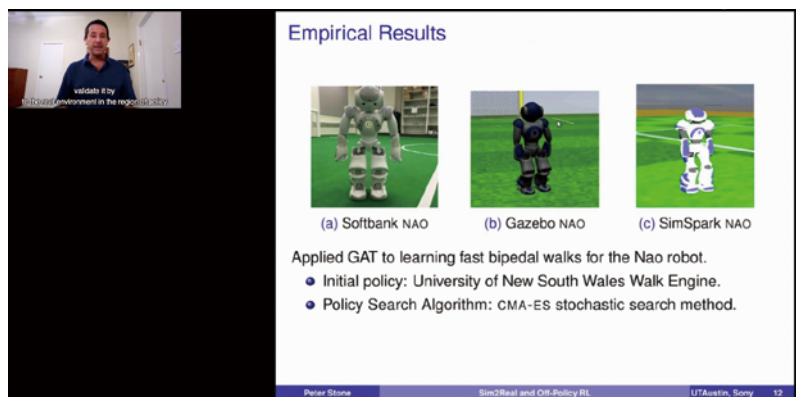
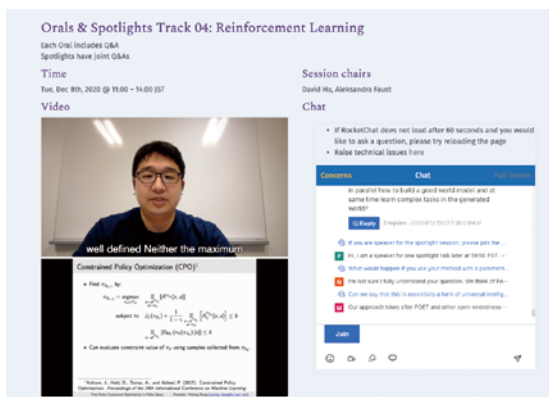
2020年12月6日から12日にかけてオンラインで開催された NeurIPS 2020に参加してきました。NeurIPSは機械学習のトップカンファレンスの1つとして数えられている国際会議であり、今年度はカナダのバンクーバーでの開催が予定されていましたが、COVID-19の世界中での流行の影響もあり、オンラインでの開催となりました。そのような状況下でありながら、今年度は過去最高となる9454本の論文が投稿され、その内の約20.1%にあたる1900本が採択されました。前年度の投稿数が6743本、採択数は1428本であったことと比較しても、機械学習分野の規模の拡大が依然として続いていることが窺えるかと思えます。オンラインでの開催ではありましたが、参加者同士の交流を促すような工夫も随所にみられました。本記事では、そんなNeurIPS2020の様子を報告したいと思います。

NeurIPS2020は事前に録画された動画と、著者とのライブセッションを交えた形式でプログラムが進行していきました。オーラルやスポットライトとして採択された論文は、プログラムで決められた時間に事前に録画された動画が中継され、その後チャットとZoom上でのリアルタイムの質疑応答が行われていました。ポスター発表はGatherTownというレトロゲームのようなドット絵の仮想空間上でのオンライン交流ツールで行われました。GatherTownでは、仮想空間上で近くにいる参加者と気軽にビデオチャットできる機能があり、これを利用して著者と直接会話することができます。ゲームのように直感的に操作することができる上、仮想空間上の所々に遊び心が見られ、とても良い体験でした。

本会議で発表された論文の中で特に注目したいのが、オフライン強化学習と呼ばれる分野の研究です。オフライン強化学習は、通常の強化学習で行われるような環境との相互作用を一切許さず、事前に収集されたデータセットのみを用い

て強化学習を行う問題設定の研究です。従来の強化学習では学習に大量の相互作用を必要とし、現実世界でのロボットなどへの応用には向かないといった欠点が指摘されていましたが、それを克服する可能性があるものとして近年盛んに議論されている分野です。今回のNeurIPS2020でも複数の研究が発表されており、一部を挙げると、固定のデータセットを用いることで生じるDistributional Shiftの影響を緩和する“Conservative Q-Learning for Offline Reinforcement Learning”がUC Berkeleyから、モデルベースオフライン強化学習で生じるダイナミクスモデルの不確実性を考慮して学習する“MOPO: Model-based Offline Policy Optimization”がStanford Universityから、アルゴリズムを評価するためのデータセットを作成した“RL Unplugged: A Collection of Benchmarks for Offline Reinforcement Learning”がDeepMindからそれぞれ発表されていました。また、UC BerkeleyのSergey LevinとAviral Kumarらによるオフライン強化学習のチュートリアルが企画されていたり、Offline Reinforcement Learningのワークショップが初めて開催されるなど、改めて分野の注目度の高さと盛り上がりを肌で感じることができました。

そのOffline Reinforcement LearningワークショップとDeep Reinforcement Learningワークショップに、筆者自身も論文を投稿し、ポスター発表を行ってきました。これまで、オンライン上でのポスター発表経験がなかったため、少々緊張と不安を感じながらの参加でしたが、特に違和感なく他の研究者と交流ができ、非常に良い刺激を受けました。来年以降の国際学会の現地開催の目処が未だ立たない中ではありますが、オンライン開催であっても積極的に参加していきたいと思いました。



協賛イベント:第28回脳の世紀シンポジウム「依存症と脳」

小林 七彩 (東京医科歯科大学 医学部 精神行動医科学分野 大学院生)

2020年9月16日に「脳の世紀シンポジウム『依存症と脳』」がオンラインで開催されました。様々な専門家の視点からみた依存症の現状と今後について、Web上で誰もが無料で聴講できるという大変貴重な機会でした。

津本忠治先生の開会の言葉にはじまり、おおたわ史絵先生の衝撃的な依存症患者の家族としての体験談、依存症研究の第一人者である池田和隆先生による、病態生理や今後の治療薬に関する可能性といった研究のお話、依存症治療や啓蒙活動の第一線で活躍される松本俊彦先生による孤独の病としての依存症患者の臨床像、昨今話題となっている発達障害とスマホとの関連について、発達障害の分野で活躍される神尾陽子先生、大隅典子先生のお話、脳画像でみるギャンブル障害についての高橋英彦先生のお話、座談会のなかではATRの川人光男先生による電気通信会社との共同研究やニューロフィードバック研究のお話もあり、最後は樋口輝彦先生の閉会の言葉で閉幕するという、本当にご高名な先生方による、充実した内容のシンポジウムでした。

特に印象的だったのは、高校時代に著書「女医の花道」を愛読しており馴染みのあったおおたわ先生の、依存症当事者の家族としての体験談でした。お母様が鎮痛剤依存となり、体は注射痕だらけ、「やめたいのにやめられない」というその病態から、裏切られたという家族の悲しみ、憎しみは相当なものだと感じました。家族にはイネイブリングをやめなさい、と私自身も含め治療者は言いますが、暴力やより重大な犯罪に及ぶのではという懸念から、決して実践することは容易ではありません。また、そのような対応をしてもすぐに変化が訪れるわけではなく、依存症の当事者とつきあっていくには大変に忍耐と労力がかかることだと改めて実感しました。

川人先生や高橋先生から話題に上がった、電気通信会社と東京医科歯科大学が共同で行っているアプリを用いたスマートフォン使用障害の患者を対象とした臨床研究に関しては私自身も関与している課題であり話題にさせていただき嬉しく思います。家庭内暴力や数百万の課金トラブルなど、深刻な問題を抱えている方がいる一方で、親御さんが過剰に心配され、何ら社会機能障害のない方が単に長時間使用というだけで受診することもあります。コロナ感染症の流行により、インターネット、電気通信機器の需要はますます高まっている昨今、神尾先生のご講演の中でもあったように闇雲に制限するというの

は意味がなく(今回のオンラインシンポジウムを視聴するだけで5,6時間は優に超えます!)スマートフォンやインターネット、ゲームの適切な使い方、どのような状態が問題なのか、といった情報を正しく伝えていかなければなりません。

依存症は「性格の問題」「本人の意思が弱い」といった誤解や、その問題行動の多さから、仕事、家族、学習の機会、健康、社会的信頼、様々なものを失ってしまい、当事者に対し陰性感情を抱く人も少なくありません。失ったものが多いほど回復には時間がかかるため、もう少し早く医療に繋がってれば、ともしどかしい思いをすることは多々あります。

依存対象の物質の摂取、または嗜癖行動を体験しなければ病気になることはない、という因果関係ははっきりしているものの、皆が依存症になるわけではなく、その個人差から臨床や研究が難しいと感じることも多いですが、先生方が御講演を通して、支援体制や、新しい依存症治療薬、バイオマーカーの探索など、今後の依存症医療の方向性を示し、期待を抱かせるシンポジウムでした。

第9回領域会議参加記

佐藤 寛紀

(東京大学 池上高志研究室 博士課程 学生)

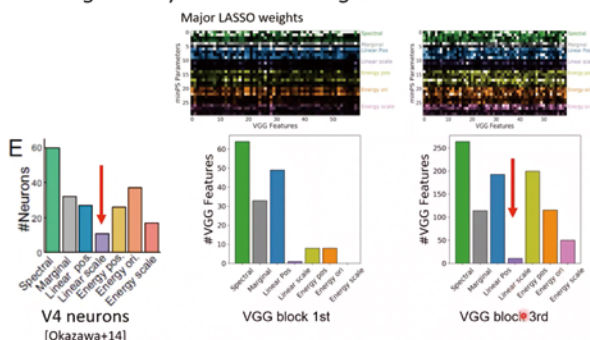
2020年10月13、14日の2日間に渡り、第9回領域会議がオンライン上で開催されました。私は新学術領域研究「人工知能と脳科学」の領域会議に参加させていただくのは今回が初めてでしたが、多くの発表・議論に刺激を受け、濃密な2日間となりました。

中でも、特に私が興味を持ったのは松尾豊先生の発表でした。その発表の中で、松尾先生は動物OSと言語アプリのアイデアについて言及され、それを基にした構想をお話しになりました。私は進化学を専門とし、知能・脳に関してはその進化的側面について関心があり、今回の会議に参加させていただいたのですが、生命がその進化の中でどのように動物OSを発展させ、言語アプリを獲得したのかは非常に興味深いテーマのように思えます。発表後、私は動物OSが存在すると仮定した場合、それは神経系上でどのように実装可能か(もしくは創発的に出現するのか)?その為にはどのようなハードが必要か?そのOSはどのようなアプリを実行可能か?そして、それらのハードとソフト(OS、アプリ)はどのように進化的に獲得しうるか?など考えを巡らせました。

本会議が開催された2020年10月には、既にコロナウイルスのパンデミック下での活動に慣れてきたという面があると思われ、オンライン開催であっても発表・議論はある程度滞りなく行われていたように思えます。しかし、コーヒー・ブレイク中の雑談のような会議らしい要素が欠け、また、全てが画面の中で行われることに起因すると思われる全体的な単調さがあったことも否めません。この原稿執筆中の2021年1月現在、パンデミックには収束の気配はなく、国内に関してはより状況の深刻度が増しているという現状を鑑みると、会議・学会のオンライン上での開催は今後もしばらく続きそうです。本会議の内容とは離れてしましますが、オンライン会議・学会の手法開発は、活発な研究活動を維持する為には急務のように感じられました。

Results: LASSO regression results

Weight analysis of LASSO regression



丸山 透

(東京大学 池上高志研究室 博士課程 学生)

元々美術大学の出身だった私にとって、博士課程1年目でもあった昨年は入学と同時にCovid-19による大学のオンライン化が始まり、ガイダンスや研究室でのコミュニケーションなどの少ない厳しい環境でのスタートとなりました。この一年は新しい分野の知見や作法などを学ぶ期間となり、そのような環境の中で、学会のオンライン化によって参加のハードルが下がったことは、(発表、質疑などに慣れる、と言う機会は逸したものの)幅広い知見を得る良き機会となったと思っております。

特に本領域会議は、人工知能と脳科学、と言う門外漢からすればとても近い学問領域のように思えますが、改めて発表を拝聴させていただくと、それぞれの分野は大きく異なることに重点を置いていることがわかってきました。言うなれば、音楽と音響学のように。

私自身の興味としては、表象の世界で創作活動をしていることもあり、表象設置問題に関する興味が強く、庄野先生の「視覚心理に基づくテキスト特徴表現と深層特徴表現のマッピング」、また、研究室内で時間生成に関する研究を分担している都合、小松先生の「予測の神経基盤: 広域皮質脳波における時空間構造」の二つの発表が強く印象に残っています。

内容に関しては、「ひとつわからない単語を調べると、5つくらいわからない単語が出てくる」というような自身の状況を鑑み、深く触れることはあえて避けようと思いますが、ともあれオンラインでのこのような会議は、ディスカッションや複数人でのコミュニケーションの難しさ、と言うものはあるものの、情報へのアクセスのしやすさ、調べながら聞くことの容易さなどの良い側面もあるように思います。

特に研究というものの性質上、様々な国、地域に住む方々の知見がFace to faceで集まる、という部分が現在の世界の情勢では難しくなっている状況で、これからどのようなオンラインでの会議がありうるのか、ということは、自身も微力ながら考えていきたいです。



第8回神経法学研究会

谷口 忠大（立命館大学 教授）

人の近代社会は法律によって整えられて、教育活動も経済活動も法の支配の下で行われています。私たちはそういう世界に住んでいるわけです。人工知能の視点から考えれば、言語を学習しその意味を理解する人工知能やロボットを作ることすら難しいのに、法律を作りそれを社会（マルチエージェントシステム）全体で守っていく人工知能を作るなんてチャレンジングすぎる問題です。さて、そこまでの話にはなかなか行かないものの、人工知能や神経科学と法学の接点は多様に存在し始めています。人工知能と脳科学（神経科学）の融合を行うのが本新学術領域であるように、法学と神経科学の接点を対象としてその融合研究を推進する研究会があります。筆者の所属大学である立命館大学の山口直也教授（法務研究科）が主宰されている神経法学研究会に参加させていただきました。この研究会は山口先生の科研費・基盤研究(B)「わが国における神経法学の基盤的研究－法学・医学・心理学の協働－」を母体とした研究会です。その第8回研究会に、本新学術領域代表の銅谷賢治先生が話題提供者として招待され、連れ立って私も参加させていただきました。同じ立命館大学の教員なので、本来ならば私が銅谷先生と山口先生をお繋ぎすべきところだったのですが、むしろ今回は銅谷先生に私と山口先生を媒介いただいた格好です。

情報系の学生などは法律とはまるでアルゴリズム的なルールとして存在しているような誤解を持ちがちですが、脳に支えられた人間の認知や意思決定と法律の世界とは切っても切れない関係にあります。神経法学では刑事法領域での刑事責任論、捜査（取調べ）段階における被疑者・参考人の心理過程、刑事裁判における目撃者の証言の信用性の問題など、人間の意思決定を基盤にする法的課題を脳科学、神経科学、認知心理学などと協働しつつ学際的に研究されています。

オンライン開催となった第8回神経法学研究会には20人を超える参加者が参加し、Zoomミーティング形式でのアットホームな雰囲気の研究会が始まりました。冒頭に山口先生から研究会の趣旨説明があり、その後、順に参加者全員が一言ずつ自己紹介をしていきました。各大学の法律系心理系の研究者のみならず、弁護士や検事といった実務家の参加者もおられました。当該新学術領域からは銅谷先生と私の他に、五味先生、疋田先生、大関先生、内部先生、松本先生などが参加されていました。それら導入の後、銅谷先生が『人工知能と脳科学の融合と社会』という題目で講演を開始。「人工知能の現在」や「脳科学と人工知能の関係性」といった点に関

して話題提供を、というリクエストに応じて、人工知能と脳科学の相互作用の歴史や、意思決定の理論である強化学習の神経基盤などが解説されました。さらに自律的な人間主体による社会を形作る法学との関係性から、そのあと特に自律性と社会性に関して自律性AIエージェントの驚異や、人間社会にAIは何を学ぶかといった話題を提供されました。特に「自律AIが危険だというのが、そもそも人間も危険な存在である。公選任期制や独占禁止法、学問においてはピアレビューなど様々な精度や法律によって乗り越えてきた。それに学ぶべきだ」というメッセージは大変示唆的でした。

質問では少年非行に関わって意思決定に関わる前頭前野の機能に関する実験結果の解釈や、自殺や犯罪の行動の抑制に関わる脳機能の計算論的解釈、虐待によって変容したかもしれない脳機能の検診に関わる質問など法律や裁判の内容に直結した学際的な質問が沢山と出ました。またより具体的に、二光子顕微鏡を用いた実験に関わる着色の方法や神経修飾物質の影響が時間スケールに関わってどう変わるのかといったテクニカルな質問もありました。常日頃から持っている脳科学や人工知能への光の当て方とはまたことなる視点に、「そういう視点か。確かにそういう問題があるのだな」と新たな気づきを与えられました。

人工知能と脳科学（神経科学）の対照と融合を目指す本新学術領域ではありますが、人工知能と脳科学の融合とその発展は、神経法学を始めとして、人間の認知に関わるあらゆる分野に接点を形成し無限の広がり成しています。新学術領域「人工知能と脳科学の対照と融合」は今年度で終了しますが、本プロジェクトに参画した研究者の中にはその広がりの中へと探索の歩みを進める研究者も多くいるでしょう。神経法学はそのようなフロンティアの一つなのかもしれません。



イベント情報

主催イベント

第10回領域会議

日程：2021.3.6-8

場所：オンライン開催

「人工知能と脳科学」

カルシウムイメージングワークショップ

日程：2021.3.20

場所：オンライン開催

脳と心のメカニズム第21回冬のワークショップ

日程：2022.1 予定

場所：北海道、ルスツリゾートホテル



AI
AND
BRAIN

発行／編集 新学術領域研究「人工知能と脳科学の対照と融合」
お問い合わせ 新学術領域研究「人工知能と脳科学の対照と融合」事務局
Mail ncus@oist.jp
2021年3月発行

www.brain-ai.jp