Gatsby-Kakenhi Workshop on AI and Neuroscience

11th and 12th May 2017

Sponsors

Gatsby Computational Neuroscience Unit, UCL
KAKENHI Project on Artificial Intelligence and Brain Science
RIKEN Center for Advanced Intelligence Project
Whole Brain Architecture Initiative
MEXT Supercomputing Project on Brain and Artificial Intelligence

Schedule 11th May 2017

9.00 – 9.30 Arrival

9.30 – 10.10 **Masashi Sugiyama (RIKEN AIP/U Tokyo)**
Title: Classification from Weak Supervision

10.10 – 10.50 **Shakir Mohammed (Deepmind)**
Title: Plausible Reasoning in Artificial Agents

10.50 – 11.30 **Daisuke Okanohara (Preferred Networks)**
Title: AI in Real World: Automobile, Robot, Bio/Healthcare, and Art creation

11.30 – 11.40 Break

11.40 – 12.20 **Arthur Gretton (Gatsby Unit)**
Title: Learning Features to Compare Distributions

12.20 – 13.00 **Jun Morimoto (ATR Computational Neuroscience Laboratories)**
Title: Motor Learning Methods for Humanoid Control

13.00 – 14.20 Lunch and Posters (See Page 16)

14.20 – 15.00 **Peter Dayan (Gatsby Unit)**
Title: Decision-Theoretic Psychiatry as Artificial Unintelligence?

15.00 – 15.40 **Hidehiko Takahashi (Kyoto University Graduate School of Medicine)**
Title: The Interface between AI and Psychiatry (Schizophrenia) Research

15.40 – 16.20 **Maneesh Sahani (Gatsby Unit)**
Title: Inference in Perception

16.20 – 16.30 Break

16.30 – 17.10 **Shinji Nishimoto (CiNet)**
Title: Linking Neuroscience and Machine Learning via Latent Features of Natural Stimuli

17.10 – 17.50 **Aapo Hyvarinen (Gatsby Unit)**
Title: Nonlinear ICA using Temporal Structure: A Principled Framework for Unsupervised Deep Learning

17.50 – 18.30 Discussion and Drinks and Nibbles
18.30 – Onwards Dinner (Brasserie)

9.00 – 9.30

Arrival

9.30 – 10.10

Masashi Sugiyama (RIKEN AIP/U Tokyo)

Classification from Weak Supervision

Machine learning from big training data is making a great success. However, there are various application domains that prohibit the use of massive labeled data. In this talk, I will introduce our recent advances in classification from weak supervision, including classification from two sets of unlabeled data, classification from positive and unlabled data, and a novel approach to semi-supervised classification.

10.10 – 10.50

Shakir Mohammed (Deepmind)

Plausible Reasoning in Artificial Agents, Shakir Mohamed, DeepMind

Our recent recent research is motivated by the goal of equipping artificial agents with the abilities needed to realistically reason and act in complex environments. These agent-abilities are varied, and include scene understanding, concept formation, future-thinking, abduction, memory, and planning, amongst many others. To be effective, each of these functions must enable plausible reasoning: the use of probabilistically consistent inferences and decision-making based on experience. This talk will explore the plausible reasoning that is possible with Bayesian deep learning and develop several agent-reasoning components: generative models with the ability for one-shot generalisation; models that can form temporally-extended future predictions; temporal models with external memory; approaches for visual concept learning; density-estimators that aid exploration; and policy search methods for hierarchical planning. Our state-of-the-art does not yet reach the ideal, common-sense agents we desire, and we will unpack some of the many remaining challenges that when overcome, will empower us to develop more realistic and practical artificial agents.

10.50 – 11.30

Daisuke Okanohara (Preferred Networks)

AI in real world: Automobile, Robot, Bio/Healthcare, and Art creation

Artificial intelligence technology based on deep learning has improved dramatically during the past few years, and new applications are continually emerging. AI is becoming practical enough that it is now being applied to solve problems in some of the world's largest industries. In this talk, I will introduce PFN's recent progress in applying AI to the following world problems: autonomous driving, industrial robotics, life science, and art creation. I will also discuss some of the challenges found in these applications.

In autonomous driving, we have developed a recognition system for outdoor environments that is accurate, fast, and robust enough for actual use cases. We have also investigated both supervised and reinforcement learning-based planning to handle complex situations. Since outdoor environments are extraordinarily diverse, it is not practical to collect data for all possible events, especially rare events. To solve this, we require new data augmentation, generation and simulation methods.

In industrial robotics, we chose the picking task as our first application because it is simple, but also challenging enough to be investigated, and huge demand exists in the market. We apply deep learning to the detection of the items, the estimation of picking angle, and contacting point. Although the current picking system can perform reasonably well for most use cases, when compared to human performance, it can still be slow and inaccurate for previously unseen items. In the industrial field, we also apply deep learning to anomaly detection, prediction, optimal control, and various other tasks of industrial importance.

In the life sciences area, we have developed a new cancer prediction model using microRNAs in collaboration with the Japanese National Cancer Center. Since MicroRNAs have rich and dynamic information in the body, we can detect various types of cancer earlier and more accurately than existing methods.

In art creation, we have released PaintsChainer (#PaintsChainer) as a demonstration service which automaticlly adds color to a line art. PaintsChainer has already painted almost million line arts. The current PaintsChainer uses GAN and U-net and we are investigating several improvements on it.

11.30 – 11.40

Break

11.40 – 12.20

Arthur Gretton (Gatsby Unit)

Learning features to compare distributions

We consider the problem of two-sample testing, where the goal is to test whether or not two samples come from the same distribution: for instance, microarray samples from cancerous or healthy tissue, or images of different types of vehicles. This is a more general setting than that of a t-test, where a difference in sample means is tested: by contrast, we consider any difference in the sample distributions (in means, variance, higher order moments, ...). A powerful test can distinguish unlike distributions with high probability. The goal of this talk is to learn features of the samples so as to maximize test power.

I will demonstrate ways of training a statistical test to maximize power, using two divergence measures: the maximum mean discrepancy (MMD), and differences of smooth features (the ME test). In both cases, the key point is that variance matters: it is not enough to have a large empirical divergence; we also need to have high confidence in the value of our divergence. Using an optimized MMD discriminator, we can detect subtle differences in the distribution of samples from a powerful generative model and real hand-written digits which humans are unable to find (for instance, small imbalances in the proportions of certain digits, or minor distortions that are implausible in normal handwriting).

12.20 – 13.00

Jun Morimoto (ATR Computational Neuroscience Laboratories)

Motor Learning Methods for Humanoid Control

We discuss how we can take advantage of the recent development of powerful computational resources to improve the policy used to control humanoid robots. We mainly introduce a hierarchical motor learning framework for humanoid robot control. Specifically, we develop a computationally efficient Model Predictive Control (MPC) method for real-time control of humanoid robots. Although MPC is a highly useful approach to deriving a policy for control of nonlinear dynamical systems, its application to a robot having many degrees of freedom is still a challenging problem because MPC is quite computationally intensive. To cope with this issue, we developed an MPC method that implements a hierarchical optimization procedure where a lower layer in the hierarchy uses a short time horizon and a small time-step size. We evaluated the proposed method on a simulated robot model.

13.00 – 14.20

Lunch (Brasserie) and Posters (Enter Gatsby 2nd floor and use spiral stairs to 3rd floor)

14.20 – 15.00

Peter Dayan (Gatsby Unit)

Decision-Theoretic Psychiatry as Artificial Unintelligence?

A central tenet of the field of computational psychiatry is to understand disorders of mentation and decision-making as apparent failure modes of the normative, ordered, operation of a collection of underlying systems. This implies that we can look to AI for clues for what can go right and wrong. We duly analyze putative failures as stemming from abnormalities in the framing of the problem or task, the mechanisms of cognition, the data available from the environment, or indeed combinations of all three of these. Along with issuing due condemnations, we consider cases in which current, or potentially evolutionarily justifiable, optima may actually be implicated.

15.00 – 15.40

Hidehiko Takahashi (Department of Psychiatry, Kyoto University Graduate School of Medicine)

The Interface between AI and Psychiatry (Schizophrenia) Research

The diagnostic criteria for psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Disease (ICD) are based on the patient's behavioral signs and symptoms. Therefore, there is an explanatory gap between phenomenological entities and neurobiological underpinnings. Recently, researchers have begun to bridge this explanatory gap using a dimensional approach. This dimensional approach is compatible with data-driven approaches. Using machine learning methods, we built a model (biomarker) to predict diagnostic labels of psychiatric disorders from resting-state functional connectivity MRI (rs-fcMRI). We developed a schizophrenia biomarker from Japanese (Kyoto) cohort, and it was successfully generalized to independent validation cohorts in the USA and Europe. We also built a prediction model of working memory capacity from (rs-fcMRI), and succeed in predicting working memory deficit in schizophrenia. Finally, if time allows, we would like to introduce our recent research to visualize altered semantic mapping in the brain of schizophrenia patients, which might be relevant to "loosening of associations", the fundamental symptoms of schizophrenia.

15.40 – 16.20

Maneesh Sahani (Gatsby Unit)

Inference in perception"

TBA

16.20 – 16.30

Break

16.30 – 17.10

Shinji Nishimoto (CiNet)

Linking neuroscience and machine learning via latent features of natural stimuli

Elucidating how the brain learns to process complex natural sensory inputs to generate appropriate behaviors is a major goal of neuroscience. Realizing this goal involves characterizing the latent neural representations at each stage of the cortical areas. Recent neurophysiological and functional imaging studies have shown that the representations learned via certain machine learning techniques closely resemble those of the cortex. Such similarities have facilitated interdisciplinary studies, including those on machine-learnt features to interpret cortical representations and cortical representations to guide machine learning. In this talk, I will introduce some of our recent studies on modeling the visual and semantic representations in the human cortex via machine learning-derived feature spaces. Our results demonstrate that learned visual and linguistic feature spaces can be used to model and decode movie-evoked cortical activities. I will also discuss how such data-driven modeling studies can facilitate the understanding of higher-order cognitive functions.

17.10 – 17.50

Aapo Hyvarinen (Gatsby Unit)

Nonlinear ICA using temporal structure: a principled framework for unsupervised deep learning

Unsupervised learning, in particular learning general nonlinear representations, is one of the deepest problems in machine learning. Estimating latent quantities in a generative model provides a principled framework, and has been successfully used in the linear case, e.g. with independent component analysis (ICA) and sparse coding. However, extending ICA to the nonlinear case has proven to be extremely difficult: A straight-forward extension is unidentifiable, i.e. it is not possible to recover those latent components that actually generated the data. Here, we show that this problem can be solved by using temporal structure. We formulate two generative models in which the data is an arbitrary but invertible nonlinear transformation of time series (components) which are statistically independent of each other. Drawing from the theory of linear ICA, we formulate two distinct classes of temporal structure of the components which enable identification, i.e. recovery of the original independent components. We show that in both cases, the actual learning can be performed by ordinary neural network training where only the input is defined in an unconventional manner, making software implementations trivial. We can rigorously prove that after such training, the units in the last hidden layer will give the original independent components. [With Hiroshi Morioka, published at NIPS2016 and AISTATS2017.]

17.50 – 18.30

Discussion and Drinks and Nibbles

18.30 – Onwards

Dinner (Brasserie)

Schedule 12th May 2017

9.00 – 9.30 Arrival

9.30 – 10.10 **Kenji Doya (OIST)**
Title: Multiple Representations and Algorithms for Learning and Behaviors

10.10 – 10.50 **Matthew Botvinick (Deep Mind)**
Title: What How Should/Can We Further Learn from the Brain?

10.50 – 11.30 **Masamichi Sakagami (Tamagawa University)**
Title: Categorical Coding of Stimulus and Inference of the Value in the Monkey Lateral Prefrontal Cortex

11.30 – 11.40 Break

11.40 – 12.20 **Zhaoping Li (Gatsby Unit)**
Title: Feedforward and Feedback Processes for Visual Detection and Recognition in Humans

12.20 – 13.00 **Hiroyuki Nakahara (RIKEN BSI)**
Title: Learning to Make Reward-Guided Decisions: Sequential, Successive, and Social

13.00 – 14.20 Lunch and Posters (See Page 16)

14.20 – 15.00 **Shane Legg (DeepMind)**
Title: Towards General Human-Like Intelligence?

15.00 – 15.40 **Tadahiro Taniguchi (Ritsumeikan University)**
Title: Symbol Emergence in Robotics: Language Acquisition via Real-World Sensorimotor Information

15.40 – 15.50 Break

15.50 – 16.30 **Karl Friston (UCL)**
Title: Active Inference and Artificial Curiosity

16.30 – 17.10 **Hiroshi Yamakawa (Whole Brain Architecture Initiative)**
Title: Brain-Inspired AI as a Way to Desired General Intelligence

17.10 – 18.00 Discussion

18.00 – 19.00 Discussion continued and Beer Hour (Brasserie)
19.00 - Onwards Dinner (Busaba Bloomsbury)

9.00 – 9.30

Arrival

9.30 – 10.10

Kenji Doya (OIST)

Multiple Representations and Algorithms for Learning and Behaviors

The brain can be seen as a multi-agent system composed of heterogeneous learners using different representations and algorithms. Different areas of the cerebral cortex provide different sensory-motor representations, based on unsupervised learning and Bayesian inference. Different parts of the cerebellum and the basal ganglia are topographically connected with the cortex and perform supervised and reinforcement learning, respectively, utilizing the representations provided by the cortex.

It should be noted that for the same behavior, multiple redundant representations can be utilized; such as allocentric, egocentric, and proprioceptive representations. Algorithms for action selection can be model-based deep search, model-free choice, or playback of a pre-learned policy or action sequence. Which of these can be most helpful can depend on the degree of prior experience, the volatility of the environment, and real-time performance constraints.

Animals and humans appear to be able to utilize these possible representations and algorithms in highly flexible ways (although we do make terrible mistakes sometimes). How such flexible selection and combination of relevant modules for a given situation can be realized is a major open problem in neuroscience. Its solution would help development of more flexible, general artificial intelligence.

10.10 – 10.50

Matthew Botvinick (Deep Mind)

What How should/can we further learn from the brain?

I'll offer some reflections on how neuroscience is most likely to be of continued use to AI, using a recent research project as a case study. The default assumption, I think, is that AI should hand to neuroscience a set of big questions, to which neuroscience can then return big answers. While this may sometimes be possible, the exchange between neuroscience and AI has historically occurred on a finer, more intimate scale, involving frequent (and often half-conscious) traversals of the disciplinary boundary. In order to foster this dynamic going forward, we need to think not only about big questions, but also about creating the right research environment. In addition to making these points, and as a way of doing so, I will describe work my group at DeepMind has been doing on the topic of meta-reinforcement learning, a project that has criss-crossed from AI to neuroscience and back again in a variety of ways.

10.50 – 11.30

Masamichi Sakagami (Tamagawa University)

<u>Categorical Coding of Stimulus and Inference of the Value in the Monkey Lateral Prefrontal Cortex</u>

In a complex and uncertain world, how do we select appropriate behavior? We believe that the brain affords at least two distinct mechanisms in learning the reward value of a stimulus: (1) Reinforcement learning through repeated stimulus-reward pairings, and (2) inferential strategy based on knowledge of the task at hand. In the current study using a reward inference task, we observed that both lateral prefrontal cortical (LPFC) and striatal neurons were able to predict reward value for a group of highly familiar stimuli (well-experienced with reward by the monkeys). But more importantly, LPFC neurons were able to infer the reward value of a new stimulus that the monkeys had not yet associated with reward directly through the integration of independently acquired stimulus-stimulus and stimulus-reward associations. In contrast, striatal neurons did not. Only after directly experiencing the new stimulus-reward contingency, were the striatal neurons able to predict reward value. Our results suggest dissociable functions in their reward prediction, i.e., that the LPFC utilizes stimulus categorization in a generative process of reward inference, whereas the striatum applies direct experiences of stimulus-reward associations in the guidance of behavior.

11.30 – 11.40

<u>Break</u>

11.40 – 12.20

Zhaoping Li (Gatsby Unit)

feedforward and feedback processes for visual detection and recognition in humans

in AI works on visual recognition, a popular method is to use a feedforward convolutional neural network for. In biological vision research, although it is known that different visual cortical areas along the visual pathway are extensively connected by both feedforward and feedback neural fibers, the knowledge about these neural connections and about their function in neural recognition is very limited. I present my recent work to reveal the feedforward and feedback, especially the feedback, processes in human visual recognition, using computationally motivated visual sychophysics methods. The experimental findings help us to infer or predict the neural architecture for the feedback processes, which can be experimentally tested.

12.20 – 13.00

Hiroyuki Nakahara (RIKEN BSI)

Learning to make reward-guided decisions: sequential, successive, and social

Reinforcement learning becomes a major paradigm for understanding neural computations and underpinnings to learn and make decisions. The research domain has also been expanding, as the field progresses. I will present our efforts aiming to contribute to the expansions, slightly in a condensed manner, given my pleasure to participate in this meeting with prominent scientists from such diverse fields. Specifically, I will highlight key features of neural reinforcement learning in sequential, successive (or contextual), and social setting, respectively: cortico-basal ganglia parallel circuits for motor sequence learning, dopamine prediction error characteristics beyond model-free learning cases, and cortical signals for learning to simulate others' decisions and making decisions with social rewards.

13.00 – 14.20

Lunch (Brasserie) and Posters (Enter Gatsby 2nd floor and use spiral stairs to 3rd floor)

14.20 – 15.00

Shane Legg (DeepMind)

Towards General Human-Like Intelligence? - Shane Legg, DeepMind

A number of major challenges in Artificial Intelligence have been solved in recent years, in some cases problems that many people did not expect to be solved for another 10 or 20 years. Does this mean that we're finally making headway towards a general human-like artificial intelligence? What would this even mean? In this talk I will offer a perspective on what intelligence is, an argument for why at least one important factor is falling into place, and a conjecture that the work that remains may be less than what you might expect.

15.00 – 15.40

Tadahiro Taniguchi (Ritsumeikan University)

Symbol Emergence in Robotics: Language Acquisition via Real-world Sensorimotor Information

This lecture provides an overview of symbol emergence in robotics. Symbol emergence in robotics is the research field that aims to create an autonomous robot that can acquire language and high-level cognitive capabilities from real-world, i.e., embodied, sensorimotor interaction with its environment. The study also aims to provide appropriate computational models that can describe human mental development. We have been developing machine learning methods that enable a robot to learn words automatically. In this talk, I am introducing several unsupervised machine learning methods.The first is for simultaneous learning of lexicons and object categories using multimodal latent Dirichlet allocation (MLDA) and nested Pitman-Yor language model (NPYLM). The second is nonparametric Bayesian double articulation analyzer (NPB-DAA) for learning phonemes and words directly from speech signals using hierarchical Dirichlet process hidden language model (HDP-HLM). The third is spatial concept acquisition and simultaneous localization and mapping (SpCoSLAM). I'm also talking about future challenges in the field and future integration of deep learning methods.

15.40 – 15.50

Break

15.50 – 16.30

Karl Friston (UCL)

Active inference and artificial curiosity

This talk offers a formal account of insight and learning in terms of active (Bayesian) inference. It deals with the dual problem of inferring states of the world and learning its statistical structure. In contrast to current trends in machine learning, we focus on how agents learn from a small number of ambiguous outcomes to form insights. I will present simulations of abstract rule-learning and approximate Bayesian inference to show that minimising (expected) free energy leads to active sampling of novel contingencies. This epistemic, curious, novelty-seeking behaviour closes 'explanatory gaps' in knowledge about the causal structure of the world, in addition to resolving uncertainty about states of the known world. We then move from inference to model selection or structure learning to show how abductive processes emerge when agents test plausible hypotheses about symmetries in their generative models. The ensuing Bayesian model reduction evokes mechanisms that have all the hallmarks of 'aha moments'.

16.30 – 17.10

Hiroshi Yamakawa (Whole Brain Architecture Initiative)

Brain-inspired AI as a way to desired general intelligence

It is critical to give artificial intelligence (AI) a general purpose problem-solving ability, as artificial general intelligence (AGI) with such functionality will bring about unprecedented intelligence development by its self-improvement capability. With the advent of deep learning, Moravec’s paradox began to be eliminated and the current AI is progressing in a bottom-up approach following the ways of phylogeny and ontogeny. However, it is still struggling with intuitive physics that can be grasped by less-than-year-old infants. From this viewpoint, the realization of AGI seems far away.

Meanwhile, from the viewpoint of AI drawing on the brain, the realization of AGI does not seem so far away, as brain functions have been partially realized with artificial neural networks (ANNs). For example, general object recognition with a convolutional neural network (CNN) can be used to model the visual temporal lobe pathway and the delayed reward calculation of reinforcement learning for the basal ganglia. Similar cases can be made for other organs such as the auditory cortex and cerebellum. Those functional models will be gradually integrated to the whole brain architecture with mesoscopic connectomic information.

As the realization of brain-inspired AGI becomes more realistic, we must have a broad perspective in research and development, for its impact on society will be extensive.

Even if AGI goes beyond our intelligence, it will be relatively easy to understand it if it operates on the same architecture as our brain. AGI based on the brain architecture will more likely be a common property of mankind, because the architecture can be agreed upon by many and be used in a widely shared development platform. Thus, in order to build AGI in harmony with human beings, it would be desirable to strengthen the cooperation between the neuroscientific community and AI community and to promote the open development of brain-inspired AI.

17.10 – 18.00

Discussion

18.00 – 19.00

Discussion continued and Beer Hour (Brasserie)

19.00 - Onwards

Dinner (Busaba Bloomsbury)

Poster presentations during both lunch sessions

Tadashi Kozuno

Data-Efficient Deep Q Learning via Action Gaps

Yukiko Matsumoto

Exploring Abnormal Semantic Representation in Schizophrenia Brain with an Encoding Model of Higher Visual Cortex

Koichi Takahashi

BriCA: A Generic Software Platform for Brain-inspired Cognitive Computing

Tadashi Yamazaki

Realtime Simulation of a Cat-scale Artificial Cerebellum on Supercomputer Shoubu

Angus Chadwick

Learning-related reorganisation of cortical inhibition

Carlos Stein

Plasticity of neural adaptation induces optimal temporal coding in cortical networks

Hiroshi Morioka

Unsupervised feature extraction by time-contrastive learning from resting-state fMRI data

Dougal Sutherland

Generative Models and Model Criticism via Optimized Maximum Mean Discrepancy