ISSN 2186-7437

NII Shonan Meeting Report

No. 137

From natural to artificial embodied intelligence: is Deep Learning the solution?

Lorenzo Jamone
Beata Grzyb
Tetsuya Ogata

November 11–15, 2019

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

From natural to artificial embodied intelligence: is Deep Learning the solution?

Organizers:
Lorenzo Jamone (Queen Mary University London, United Kingdom)
Beata Grzyb (University College London, United Kingdom)
Tetsuya Ogata (Waseda University, Japan)

November 11–15, 2019

Overview of the Meeting

Artificial Embodied Intelligence is the intelligence of an artificial agent that has a physical body, e.g. a robot. Creating such intelligence is the primary focus of Cognitive Robotics.
More specifically, Cognitive Robotics has the twofold objective of i) creating useful robots by taking inspiration from biology, and ii) obtaining a better understanding of biology by using robots as scientific tools. In most cases, the aspect of *biology* that is under consideration in this context is *human intelligence*. This is motivated by the fact that human intelligence i) allows humans to do the amazing things they do (and that current robots cannot do), and ii) is still only partially understood (there's a lot more to understand).

What is evident is that humans possess remarkable cognitive skills, acquired progressively over time through development and learning, that allow them to successfully interact with the physical environment and engage in social interactions. Researchers have been attempting to model such cognitive capabilities in artificial systems for a long time now, first in computers [1, 2] and more recently in robots as well [3, 4].

Interestingly, the recent theories of embodied cognition [5] and grounded cognition [6] have highlighted how the presence of a body, endowed with action and perception capabilities, is crucial for intelligence to emerge. In this light, robots have become even more interesting tools to understand human intelligence; at the same time, the positive impact that truly intelligent robots can have in our society is easy to foreseen and clearly impressive. However, although robots are becoming increasingly sophisticated, in terms of appearance, technology and skills, their interaction abilities in the real world are still very limited; this is true both for the interaction with the unstructured environment (e.g. unknown objects and spaces) and even more dramatically for the interaction with people. At the moment, robots still lack the required level of *human-like intelligence* that would make these interactions effective.

Notably, the world of Artificial Intelligence has been recently transformed by the increasing success of Deep Learning algorithms in many fields of applied computation: e.g., speech processing, image processing, data mining, finance, genomics [7]. Powered by smart reinforcement learning strategies, deep neural networks are successfully playing video games exceeding human abilities [8], and they have even beaten top professional human players at Go [9].

Can Deep Learning transform (or at least, contribute to) the field of Cognitive Robotics as well? Is this happening already?

So far, Deep Learning has brought promising results in the area of robot learning, yet still in very limited settings, in particular for the learning of visuomotor mappings that would allow eye-hand coordination [10], grasping [11], manipulation [12]. Three recently published review articles have discussed some of the applications of Deep Learning to robotics [13, 14], including developmental robotics [15], highlighting both potentials and limitations.

In this Shonan meeting we extended this discussion to the area of Cognitive Robotics, and we investigated to what extent Deep Learning techniques can facilitate the creation of powerful computational models to: i) control intelligent robots and deploy them in different areas of our society, and ii) increase our understanding of the human mind.

We have gathered 13 international senior researchers from different disciplines: cognitive science, experimental and developmental psychology, robotics, artificial intelligence, machine learning, human-robot and human-machine interaction.

Following the successful experience of recent Shonan Meetings on related topics (in which some of the organisers and invited speakers of this meeting were participating), we did not have a detailed program for the five days, but we steered the discussion from a very general initial theme (artificial embodied intelligence) to converge to a smaller and more focused topic (the limits and prospects of Deep Learning for the field of Cognitive Robotics), that will be the subject of a joint publication that is currently under preparation by the participants.

We believe that the interdisciplinary discussion among a group of established researchers with such diverse background is a unique opportunity to identify the opportunities provided by the current technologies towards the goals of artificial intelligence, to stimulate new promising directions for future research, and to foster novel international collaborations with strong potential for innovation.

Overview of Talks

Through a series of interactive presentations, we discussed the role of Deep Learning in Cognitive Robotics, and whether it will be *the solution* to embodied artificial intelligence.
We eventually agreed that Deep Learning is a powerful tool that can be useful to solve (or, to advance) many of the sub-problems of artificial intelligence, but is not *the solution*: in the artistic rendition below, we summarized how, more than "Deep Learning", *the solution* is to create and maintain a "Deep" bridge that brings together different scientific disciplines (e.g. machine learning, computer science, cognitive science, robotics, psychology, neuroscience) and that facilitates "Learning" from each other.
The list of presentations follows.

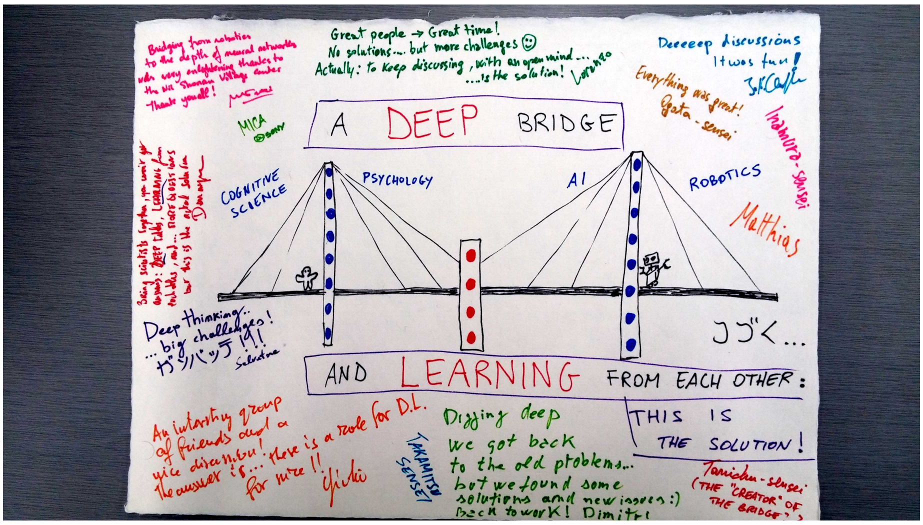

Figure 1: Artistic drawing representing *the solution* to artificial embodied intelligence.

General Robotic Intelligence: the What and Why Not.

Lorenzo Jamone, Queen Mary University of London, UK

The ability to perform everyday physical tasks in a variety of different environments is a wonderful manifestation of human intelligence. Robots cannot do that. They might be able to learn (or to be programmed for) a single task in a specific environment, but they lack the ability to generalize to new tasks and different environments. In the talk, some of the major challenges towards a General Robotic Intelligence were highlighted.

Deep predictive learning for robot intelligence.

Testuya Ogata, Waseda University

Almost a decade has passed since deep learning became the core technology of artificial intelligence. However, most of its applications are used only in cyberspace, such as image processing and natural language processing. There are very few application of systems (robots) that operate their own bodies in the "real world". Why? It is important to think about the reason when considering the essence of intelligence. In this talk, I introduce our research on real robot based on the framework of predictive learning of the sensory-motor experience, by showing the examples of joint works with multiple companies. Furthermore, it is shown that the latent space of the deep learning model acquired by this "experience-based learning" is not a continuous probability distribution structure, but a complex structure with discontinuities and fractals. I show the hypothesis that this structure contributes to the intelligence of the model.

From Humanoid Robots to Anthropomorphic Minds.

Giulio Sandini, Italian Institute of Technology

During the talk, I have argued that even if robots are motorically and sensorially very skilled and extremely clever in action execution, the technologies supporting their interaction with humans are still very primitive. I specifically addressed three points. First the fact that the asymmetry between action execution and understanding is rooted in our limited knowledge of the mechanisms at the basis of human social interaction and in particular in our ability to anticipate our own actions and those of others. Second that discovering the principles of mutual understanding is a necessary intermediate step to investigate alternative "artificial" technologies implementing such principles. Finally I stressed the fact that robotics can serve a very crucial role by joining forces with the communities studying embodied intelligence and the cognitive aspects of social interaction.

Symbol emergence in robotics: towards developmental artificial embodied intelligence.

Tadahiro Taniguchi, Ritsumeikan University

Symbol emergence in robotics aims to develop a robot that can adapt to the real-world environment, human linguistic communications, and acquire language from sensorimotor information alone, i.e., in an unsupervised manner. This line of studies is essential not only for creating a robot that can collaborate with people through human-robot interactions but also for understanding human cognitive development. This talk introduces the recent development of integrative probabilistic generative models for language learning, e.g., spatial concept formation with simultaneous localization and mapping, and vision of symbol emergence in robotics. I will also introduce challenges related to the integration of probabilistic generative models and deep learning towards developmental artificial embodied intelligence.

Adaptive Social Perception.

Dimitri Ognibene, University of Essex

Unstructured social environments, e.g. building sites, release an overwhelming amount of information yet behaviorally relevant variables may be not directly accessible. Currently proposed solutions for specific tasks, e.g. autonomous cars, usually employ over redundant, expensive, and computationally demanding sensory systems that attempt to cover the wide set of sensing conditions which the system may have to deal with. Adaptive control of the sensors and of the perception process input is a key solution found by nature to cope with such problems, as shown by the foveal anatomy of the eye and its high mobility and control accuracy. The design principles of systems that adaptively find and selects relevant information are important for both Robotics and Cognitive Neuroscience. At the same time, collaborative robotics has recently progressed to human-robot interaction in real manufacturing. Measuring and modeling task specific gaze behaviours is mandatory to support smooth human robot interaction. Indeed, anticipatory control for human-in-the-loop architectures, which can enable robots to proactively collaborate with humans, heavily relies on observed gaze and actions patterns of their human partners. The talk will describe several systems employing adaptive vision to support robot behavior and their collaboration with humans. These systems also provide insights on the developmental and computational processes behind the execution of these functions in humans.

Reinforcement Learning for Real-world Robot Control.

Takamitsu Matsubara, Nara Institute of Science and Technology

Reinforcement learning has been popular nowadays; however, its successful applications to real-world robots are still not many due to several difficulties. In this talk, I introduce our recent challenges on the application of reinforcement learning to real-world robots, such as robotic cloth manipulation, boat autopilot, and garbage crane control.

Intrinsically Motivated Curriculum Learning with Hierarchical Multi-Timescale Reinforcement Learning And Active Imitation Learning.

Sao Mai Nguyen, IMT Atlantique

Intrinsic motivation heuristic, and in particular competence progress, can be used a common criteria for a robot to choose its learning curriculum. For imitation learning, it can learn by trial and error : what and whom to imitate. In multi-task learning, it can also learn what to learn at each step of its exploration : by generating control tasks and by goal babbling. The control tasks can be simple or composite tasks needing primitive actions or a sequence of primitive actions. Using intrinsic motivation and robot planning, it can learn to associate for each parametrized task, a priori unbounded sequences of primitive actions, by discovering the dependencies between tasks. For hierarchical Reinforcement Learning, our methods leverage time-abstract representations of sequences of actions : procedures or planning. These results were shown both on robot arms manipulation and mobile robot navigation tasks.

Self-Organized Multi-Level Working Memories Facilitate Predictive Coding Based Action Planning.

Jeffrey Queisser, Okinawa Institute of Science and Technology

In contrast to the vast majority of studies on working memory in a machine learning context, the focus of the current work is on self-organization and the manipulation of working memory content. The core idea is based on a learning system that does not predict future states of the environment and its own actions directly, it rather learns a manipulation sequence of an initial state of the environment. By learning the manipulation and not the sensory predictions directly, an improved generalization of the system to new situations is expected. The presented evaluation is performed on multi-modal toy data and off-line recordings of real robot actions. Evaluation metrics include the prediction error and the success rate of motor plans for previously unseen tasks. The expected contribution to the scientific community of the current work is threefold: 1) improvement of the generalization capabilities of predictive coding based planning tasks; 2) analysis of self-organized higher-level representations and their correlation to high-level task states; 3) review of functional correlations of working memory manipulation in the brain and in the proposed artificial neural network.

From social interaction to ethical AI: a developmental roadmap.

Matthias Rolf, Oxford Brookes University

AI and robot ethics have recently gained a lot of attention because adaptive machines are increasingly involved in ethically sensitive scenarios and cause incidents of public outcry. Much of the debate has been focused on achieving highest moral standards in handling ethical dilemmas on which not even humans can agree, which indicates that the wrong questions are being asked. While traditionally engineered artifacts, including AI, require the designer to ensure ethical compliance, learning machines that change through interaction with people after their deployment can not be vetted in just the same way. I will argue that in order to progress on this issue, we need to look at it strictly through the lens of what behavior seems socially acceptable, rather than idealistically ethical. Machines would then need to determine what behavior is compliant with social and moral norms, and therefore be receptive to social feedback from people. I will discuss a roadmap of computational and experimental questions to address the development of socially acceptable machines, and emphasize the need for social reward mechanisms and learning architectures that integrate these while reaching beyond limitations of traditional reinforcement-learning agents. I suggest to use the metaphor of "needs" to bridge rewards and higher level abstractions such as goals for both communication and action generation in a social context. We then suggest a series of experimental questions and possible platforms and paradigms to guide future research in the area.

How to evaluate the cognitive function in Human-Robot Interaction?

Testunari Inamura, National Institute of Informatics

The standard approach to evaluate the performance of human-robot interaction (HRI) is a subjective evaluation, for example, using questionnaires. Because such subjective evaluation is time-consuming, an alternative evaluation method based on only objective factors (i.e. human reaction behavior) is required for real-time learning by robots to improve the quality of interaction ability. This talk introduces a research challenge to investigate the extent to which subjective evaluation results can be approximated using objective factors. Since the investigation requires a lot of human-robot interaction dataset, a virtual reality platform to progress the data collection is proposed. A robot competition framework to observe and store the HRI history data is also proposed. The robot must generate comprehensible and unambiguous natural language expressions and gestures to guide inexpert users in virtual everyday environments. By comparing the interaction history and subjective evaluation results by third-parties, it is revealed that the subjective evaluation criteria for HRI could be approximated with objective evaluation such as the embodied physical movement of test users in VR. It shows the effectiveness of the proposed VR interaction system and the possibility of real-time evaluation of interaction ability by the robot to improve the cognitive interaction skill between humans.

Social Artificial Agents and Neuro-Developmental Deficits.

Salvatore Anzalone, University of Paris 8

Endowing agents with socio-cognitive skills translates mainly into two challenges: the analysis of the dynamics of social interactions between humans; the development of skills that explicitly take in account the human presence in the agent's perception-cognition-action loop. In this presentation, I will show how a social agent can deal with people's mental states as well with its owns. This will be achieved through the exploitation of basic socio-cognitive skills as engagement, imitation, joint attention and perspective taking, together with a fine characterization and modeling of humans' individual differences. In the last decade, social agents were employed in the support of children with Neuro-Developmental Deficits, where social skills are impaired. In this case, social agents can assure a continuous, intensive, long-term caring of children. I will focus, in particular, on their use at school: an environment in which social agents can be used as tools to tailor teaching strategies and intensives therapies to the special needs of children.

What about Deep Learning methods for Assistive Technologies and Accessibility?

Dominique Archambault, University of Paris 8

Our team focuses on applied research on Assistive Technologies for persons with disabilities and accessibility in all its forms. Assistive technology (AT) refers to all kinds of technology that may help a person in a situation when a disability prevents this person from completing some tasks. For instance, the wheelchair make it possible for a person whose legs are not functioning to go from a place to another. Accessibility can be defined, in a practical way, as all means that make environment compatible with AT, that is allowing anybody, whatever technology is used, to fully participate. Our team works especially on different categories of AT: access to information, mobility of people with visual impairment, sign language dictionary, reading and writing help for people with specific learning disabilities, social robotics for helping children with neurodevelopmental troubles. The talk questions the use of deep learning methods in our field and especially in the field of social robotics for assistance. Indeed in more and more AT situations, the use of such methods seems promising. For instance people with dyslexia have important problems with correct writing and their errors are very different than those from mainstream people, making mainstream spell-checkers barely useful. DL methods may be a good approach as we have large text databases of people with dyslexia's writings.

Emergence of language in robots.

Michael Spranger, SONY

The talk introduces problems around the emergence of natural language and meaning in robots that present significant challenges to current deep learning systems. I then discuss various solutions that address these issues specifically through systems that combine reasoning and learning in a unified framework.

List of Participants

- Lorenzo Jamone, Queen Mary University London
- Tetsuya Ogata, Waseda University
- Giulio Sandini, Istituto Italiano di Tecnologia
- Tadahiro Taniguchi, Ritsumeikan University
- Dimitri Ognibene, University of Essex
- Takamitsu Matsubara, Nara Institute of Science and Technology
- Sao Mai Nguyen, IMT Atlantique
- Jeffrey Queisser, Okinawa Institute of Science and Technology
- Matthias Rolf, Oxford Brookes University
- Tetsunari Inamura, National Institute of Informatics
- Salvatore Anzalone, University of Paris 8
- Dominique Archambault, University of Paris 8
- Michael Spranger, Sony Computer Science Laboratories Inc.

Figure 2: Group picture of participants.

Meeting Schedule

Check-in Day: November 10 (Sun)

- Welcome Reception

Day1: November 11 (Mon)

- Start of Meeting
- General introduction and self-introductions of participants.
- Lunch
- Definition of the objectives of this meeting; plan for final outcome of the meeting, i.e. to write a joint survey paper based on the discussion.
- Dinner

Day2: November 12 (Tue)

- Short talks by participant #1 to #7, with focused QA.
- Group picture!
- Lunch
- Short talks by participant #8 to #13, with focused QA.
- Dinner

Day3: November 13 (Wed)

- Definition of the structure of the joint survey paper; discussion on what scientific questions to address.
- Lunch
- Excursion to Jomyoji temple, including Japanese Tea ceremony, and Social Dinner

Day4: November 14 (Thu)

- Preparation of the draft of the joint survey paper, and related discussion.
- Lunch
- Preparation of the draft of the joint survey paper, and related discussion.
- Dinner

Day5: November 15 (Fri)

- Final Discussion and Plans for Continuation of Work
- Wrap up
- Lunch
- End of Meeting

References

[1] A. Newell, H. Simon, Computer simulation of human thinking, Science 134 (3495) (1961) 2011–2017.

[2] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall, 1995.

[3] M. Lungarella, G. Metta, R. Pfeifer, G. Sandini, Developmental robotics: a survey, Connection Science 15 (4) (2003) 151–190.

[4] M. Asada, et al., Cognitive developmental robotics: A survey, IEEE Transactions on Autonomous Mental Development 1 (1) (2009) 12–34.

[5] M. L. Anderson, Embodied cognition: A field guide, Artificial Intelligence 149 (1) (2003) 91–130.

[6] L. W. Barsalou, Grounded cognition, Annu. Rev. Psychol. (2008).

[7] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015).

[8] V. Mnih, et al., Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.

[9] D. Silver, et al., Mastering the game of go with deep neural networks and tree search, Nature 529 (2016).

[10] S. Levine, C. Finn, T. Darrell, P. Abbeel, End-to-end training of deep visuomotor policies, Journal of Machine Learning Research 17 (39) (2016) 1–40.

[11] L. Pinto, A. Gupta, Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours, in: International Conference on Robotics and Automation ICRA, 2016.

[12] P. C. Yang, K. Sasaki, K. Suzuki, K. Kase, S. Sugano, T. Ogata, Repeatable folding task by humanoid robot worker using deep learning, IEEE Robotics and Automation Letters 2 (2) (2017) 397–403.

[13] H. Pierson, M. Gashler, Deep learning in robotics: a review of recent research, Advanced Robotics 31 (16) (2017) 821–835.

[14] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, P. Corke, The limits and potentials of deep learning for robotics, The International Journal of Robotics Research 37 (4-5) (2018) 405–420.

[15] O. Sigaud, A. Droniou, Towards deep developmental learning, IEEE Transactions on Cognitive and Developmental Systems 8 (2) (2016) 99–114.