Collective Predictive Coding as Model of Science

Workshop abstract

The Collective Predictive Coding (CPC) hypothesis, proposed by Taniguchi (谷口 2020, 2024), explains the emergence of a symbolic system through communication among cognitively collaborating agents. The key insight of the CPC hypothesis is that the semantic content of symbols used by agents converges through distributed Bayesian inference. Empirical studies with artificial agents have shown that, without pre-programming, two robots exchanging sensory reports can develop common names for objects that enable them to communicate effectively (Taniguchi et al., 2022).

Recently, the same framework has been applied to explain various aspects of scientific activity, including observation, experimentation, scientific communication, paper writing, peer review, and, most notably, the establishment of scientific knowledge (Taniguchi et al., submitted). This model, called Collective Predictive Coding as a Model of Science (CPC-MS), views scientists as agents who interact with their environment in diverse ways. These scientist-agents externalize the knowledge gained from observations and experiments through papers and other forms of communication. Some of this knowledge is incorporated into global scientific understanding through sharing and scrutiny (such as peer review) with other agents. This entire process is modeled as decentralized Bayesian updating, with the estimated posterior distribution representing the body of scientific knowledge.

The CPC-MS model suggests intriguing results in relation to active inference and singular models and offers significant implications for the philosophy of science, touching on themes of objectivity, scientific progress, and automation. It also provides insights into the "science of science." In this workshop, we will introduce these implications and discuss possible future directions for development.

The workshop will be conducted in English.

References

谷口忠大 (2020)『心を知るための人工知能:認知科学としての記号創発ロボティクス』, 共立出版.

Taniguchi, T., Yoshida, Y., Taniguchi, A., & Hagiwara, Y. (2022). Emergent communication through Metropolis-Hastings naming game with deep generative models. *Advanced Robotics*, 37 (19), 1266–1282.

谷口忠大 (2024)「集合的予測符号化に基づく言語と認知のダイナミクス: 記号創発ロボティクスの新展開に向けて」, 認知科学31-1, pp. 186-204.

Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., Hamada, T. (submitted). Collective Predictive Coding as Model of Science: Formalizing Scientific Activities towards Generative Science.

Collective Predictive Coding Hypothesis and Beyond

Tadairo Taniguchi

Graduate School of Informatics, Kyoto University
Research Organization of Science and Technology, Ritsumeikan University

Humanity has accumulated and passed down various knowledge and cultures within and across communities through the formation of language and communication. How to computationally express such phenomena has been an important question in research on language evolution, symbol emergence, and emergent communication. The author has been proposing a systemic view called symbol emergence systems and has taken a constructive approach to studying symbol emergence, known as symbol emergence in robotics (Taniguchi et al. 2016, 2019). Based on these constructive studies, the Collective Predictive Coding (CPC) hypothesis was proposed (Taniguchi 2024). This CPC hypothesis is based on a model that can express the formation of internal representations as representation learning through individual predictive coding. It proposes that symbol emergence, as the formation of external representations, can be expressed as collective predictive coding among multiple agents.

The CPC hypothesis aims to provide a unified computational framework for understanding how symbolic communication, particularly language, emerges in human societies, and how internal representations are formed in human cognitive systems. The CPC hypothesis argues that symbol emergence is viewed as social representation learning, which acts as distributed Bayesian inference. This distributed inference is embodied through language games, whose representative model is the Metropolis-Hastings naming game (Taniguchi et al. 2023), where each agent makes autonomous decisions to reject or adopt signs referring to their respective beliefs. The idea has been tested in an experimental semiotics study (Okumura et al. 2023). In essence, the CPC hypothesis proposes that language emerges as a collective effort to predict and encode the sensory experiences of all members of a society. It extends the concept of predictive coding from individual brains to the societal level, suggesting that symbol systems like language arise from a decentralized process of minimizing prediction errors across a population of agents interacting with their environment and each other.

After the proposal of the CPC hypothesis, we are gradually realizing that the total structure of the CPC seems to be relevant to scientific activity. The CPC model internally embraces not only the bottom-up formation of symbol systems reflecting the world structure based on observations but also top-down constraints given to the agents who are participating in communication using symbol systems. Also, the language game, including propose and acceptance/reject decisions referring to their own beliefs, is analogous to scientific communications, e.g., discussion and submitting and reviewing papers. Such systematic correspondence between symbol emergence and scientific activities leads us to the application of CPC to scientific activities in society, i.e., CPC as a model of science (CPC-MS) (Taniguchi et al. 2024).

This presentation will introduce the CPC hypothesis and then the basics of CPC-MS as an extension of the CPC hypothesis. Also, some additional implications will be

presented.

References

1. Okumura, R., Taniguchi, T., Hagiwara, Y., & Taniguchi, A. (2023). Metropolis-Hastings algorithm in joint-attention naming game: experimental semiotics study. Frontiers in Artificial Intelligence, 6, 1235231.
2. Taniguchi, T. (2024). Collective predictive coding hypothesis: Symbol emergence as decentralized bayesian inference. Frontiers in Robotics and AI, 11, 1353870.
3. Taniguchi, T., Nagai, T., Nakamura, T., Iwahashi, N., Ogata, T., & Asoh, H. (2016). Symbol emergence in robotics: a survey. Advanced Robotics, 30(11-12), 706-728.
4. Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., & Hamada, T. (submitted). Collective Predictive Coding as Model of Science: Formalizing Scientific Activities towards Generative Science. (arXiv preprint arXiv:2409.00102, 2024)
5. Taniguchi, T., Ugur, E., Hoffmann, M., Jamone, L., Nagai, T., Rosman, B., Matsuka, T., Iwahashi, N., Oztop, E., Piater, J., & Wörgötter, F. (2019). Symbol Emergence in Cognitive Developmental Systems: A Survey. IEEE Transactions on Cognitive and Developmental Systems, 11(4), 494-516.
6. Taniguchi, T., Yoshida, Y., Matsui, Y., Le Hoang, N., Taniguchi, A., & Hagiwara, Y. (2023). Emergent communication through metropolis-hastings naming game with deep generative models. Advanced Robotics, 37(19), 1266-1282.

How Collective Predictive Coding Hints at the Future of Science with AI

Shiro Takagi

Independent Researcher

AI has made remarkable progress, its influence spreading throughout society. Science is no exception to this trend; AI has become an innovative tool in the field, and its applications in science are rapidly expanding (Wang et al., 2023). Furthermore, AI's development isn't limited to its use as a mere tool; scientists are now exploring whether AI can do research on its own as scientists (Zenil et al., 2023, Lu et al., 2024). They're trying to figure out if AI can come up with its own research questions, plan and run experiments, and make sense of the results. The possibility of AI acting as an autonomous scientist points towards a future novel scientific community where both human and AI scientists coexist and contribute to scientific endeavors (Krenn et al., 2022, Messeri and Crockett 2024). Given this rapid advancement of AI for science and its potential to revolutionize scientific practices, there's growing interest in how AI is affecting science.

Recently a new model of science Collective Predictive Coding as Model of Science (CPC-MS) was proposed (Taniguchi et al., 2024). The main feature of CPC-MS is its proposal of a model of science as a CPC activity involving multiple agents. This characteristic allows us to discuss important aspects of how AI might change the nature of science. Specifically, by modeling the scientific community as a hybrid system composed of fundamentally different agents----AI and humans----CPC-MS enables us to explore how this transformation might impact science or even alter the very nature of scientific inquiry. In this talk, I will discuss the implications from CPC-MS on how AI impacts science.

References

Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. Nature, 620(7972):47–60, 2023.

Hector Zenil, Jesper Tegn´er, Felipe S Abrah˜ao, Alexander Lavin, Vipin Kumar, Jeremy G Frey, Adrian Weller, Larisa Soldatova, Alan R Bundy, Nicholas R Jennings, et al. The future of fundamental science led by generative closed-loop artificial intelligence. arXiv preprint arXiv:2307.07522, 2023.

Chris Lu, Cong Lu, Robert Lange, Jakob N Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery. arXiv preprint arXiv:2408.06292, 2024.

Mario Krenn, Robert Pollice, Si Yue Guo, Matteo Aldeghi, Alba Cervera-Lierta, Pascal Friederich, Gabriel dos Passos Gomes, Florian H¨ase, Adrian Jinich, AkshatKumar Nigam, et al. On scientific understanding with artificial

intelligence. Nature Reviews Physics, 4(12):761–769, 2022.

Lisa Messeri and MJ Crockett. Artificial intelligence and illusions of understanding in scientific research. Nature, 627(8002):49–58, 2024.

Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., Hamada, T., Collective Predictive Coding as Model of Science: Formalizing Scientific Activities towards Generative Science. arXiv preprint arXiv:2409.00102, 2024.

On the Collective Predictive Coding Hypothesis and the Phase Transition Phenomena in Multi-Agent Systems

Yusuke Hayashi

AI Alignment Network

This presentation explores how scientific progress can be modeled as a decentralized Bayesian inference process using the Collective Predictive Coding (CPC) framework in multi-agent systems. CPC posits that agents, including scientists, engage in collective efforts to minimize prediction errors by continuously updating their internal models through communication and collaboration. These collective actions result in shared external representations of knowledge, analogous to scientific theories.

Scientific progress is framed as an optimization of posterior distributions, where individual agents refine their hypotheses based on new data and feedback from the broader community. The presentation highlights how scientific knowledge evolves through both gradual improvements—characteristic of normal science—and phase transitions, which represent paradigm shifts. These transitions occur when accumulated anomalies in existing theories lead to critical points, prompting the adoption of new models that better explain the data.

Drawing on singular learning theory, the presentation explains how these phase transitions in scientific knowledge resemble shifts in the posterior distribution from one local optimum to another. This provides a formal account of paradigm shifts, where the collective understanding undergoes rapid and fundamental changes. The presentation also discusses the generative nature of science, emphasizing that scientific knowledge not only reflects current understanding but also drives the generation of new hypotheses and research directions.

The role of collective intelligence is central to this framework, as it highlights how decentralized collaboration among agents corrects individual biases, leading to a more accurate and objective understanding of the world. The integration of AI into this process is explored, with AI potentially enhancing the diversity of perspectives in scientific discovery. However, communication challenges between human and AI agents must be addressed to fully realize this potential.

This approach offers a novel perspective on the dynamics of scientific progress, demonstrating how collective predictive coding can model both continuous developments and revolutionary shifts in scientific paradigms.

References

1. Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., Hamada, T., Collective Predictive Coding as Model of Science: Formalizing Scientific Activities towards Generative Science. arXiv preprint arXiv:2409.00102, 2024.
2. Watanabe, S., Algebraic Geometry and Statistical Learning Theory, Cambridge University Press, 2009.
3. Watanabe, S., Mathematical Theory of Bayesian Statistics, Routledge, 2020.

Collective Predictive Coding for Collective Curiosity and Exploration

Taiyo Hamada

Araya Inc.

Abstract TBA.

From Confirmation to Generation: Rethinking Science through Collective Predictive Coding

Jun Otsuka

Graduate School of Letters, Kyoto University
RIKEN Center for Advanced Intelligence Project
Data Science and AI Innovation Research Promotion Center, Shiga University

The CPC-MS framework offers a bird's-eye perspective on scientific activities and knowledge formation, while also proposing a new way of thinking about science. In particular, it encourages a shift from the traditional understanding of science centered on confirmation to one focused on generation, that is, the formation of hypotheses and predictions. This presentation discusses the differences in these views of science and their philosophical implications.

Traditional views of science and philosophy of science have emphasized the role of confirmation in science. Logic and statistics have played key roles as methodologies for hypothesis validation. Scientific knowledge is seen as the accumulation of hypotheses justified by valid confirmation methods. In contrast, CPC-MS sheds light on the more pragmatic aspects of science. Here, statistics is not so much a tool for filtering truth but rather a protocol for facilitating peer review and mutual evaluation. Scientific knowledge shared by the academic community is also characterized by its generative nature, emphasizing its role in making predictions and suggesting new research directions.

This generative view of science simultaneously underscores the social nature of the objectivity of scientific knowledge. As Longino (1990) and Kitcher (1993) have discussed, scientific objectivity is not guaranteed by individual scientists following universal methodologies, but is rather socially constructed through the mutual dialogue and criticism of researchers with diverse methods and motivations. In the CPC-MS framework, the posterior distribution that represents scientific knowledge is approximated through the sampling process from the scientific community. The diversity of the scientific community can be interpreted as a condition for ensuring that this Markov chain converges to the correct posterior distribution.

Additionally, by understanding scientific progress as an improvement in predictive accuracy, CPC-MS offers a potential solution to Kuhn's problem of incommensurability between paradigms. A paradigm shift can be seen as a significant reconfiguration or "jump" in the posterior distribution. While the hypotheses accepted by the scientific community change dramatically before and after such jumps, resulting in a certain level of incommensurability, it remains possible to compare them from the standpoint of prediction errors derived from the posterior distribution.

This presentation will examine these topics and explore the philosophical implications of the CPC-MS framework.

References

Philip Kitcher. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press, 1993.

Helen E Longino. *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press, 1990.