People of Data
Zooming in on the brain via data science

Shanshan Jia[1,2] and Zhaofei Yu[1,2,*]
[1]Peking University, Institute for Artificial Intelligence, Beijing 100871, China
[2]Peking University, School of Computer Science, Beijing 100871, China
*Correspondence: yuzf12@pku.edu.cn
https://doi.org/10.1016/j.patter.2022.100463

Zhaofei Yu, a young principal investigator (PI), and Shanshan Jia, a PhD student, explain the importance of data science in their research. They have developed a method to extract information from high-dimensional data based on wavelet decompositions using simulated and experimental neuroscience data. This method is not only beneficial for neuroscientists, but it might also be helpful to anyone who deals with high-dimensional data.

Zhaofei and Shanshan, tell us about your background (personal and/or professional)

Zhaofei Yu: I received my PhD in control science and engineering from Tsinghua University, with a focus primarily in computational neuroscience, machine learning, and big data. After that, I was a postdoctoral fellow at Peking University, where I focused on neuromorphic computation. Additionally, I was a visiting scholar at Graz University of Technology from 2015 to 2016, where I worked with Wolfgang Maass on spiking neural networks. Currently, I am an assistant professor at Peking University. My current research interests include artificial intelligence (AI), brain-inspired computing, and computational neuroscience.

Shanshan Jia: My main research direction is brain-inspired computation. Under the guidance of my tutors, I began to study the relationship between stimulation and neural response. On the one hand, we infer the internal computing components of the neural network according to stimulation and response. On the other hand, we aim to mine meaningful information from high-dimensional data such as brain data, especially the patterns containing external stimuli and/or conditions.

Zhaofei, tell us about the research in your team: What drew you to this area of research? How has the research focus of your team evolved over the years?

ZY: The research in my team includes two aspects: (1) AI for science and (2) science for AI. On the one hand, we use AI tools, like deep neural networks, to extract meaningful information embedded in high-dimensional neuroscience data and explain how the visual system implements encoding and decoding. On the other hand, we build brain-like visual models to solve computer vision's bottleneck problem. This is mainly related to my research experience. I focused on AI for science during my PhD, and science for AI during the post-doctoral period. I decided to combine them when I set up my team.

Zhaofei, tell us about your paper in this issue of *Patterns*

ZY: Our work addresses a crucial concept in data science: extracting meaningful information embedded in high-dimensional data.[1] We leveraged conditional mutual information between wavelets to select a small subset of non-redundant features and demonstrated the efficiency and effectiveness of these features by decoding the stimulus or condition with high accuracy on various types of neuroscience data. Our work may have wide implications for other types of data with rich temporal dynamics.

Zhaofei, why did you decide to publish in *Patterns*?

ZY: *Patterns* publishes high-quality and groundbreaking original data science research. Our work of representing the dynamics of high-dimensional data enables the novel model design of machine learning with representative features. Thus, they will be of great interest to readers of *Patterns*. Moreover, our work fits the aim of *Patterns*, which is to share data science solutions for problems that cross domain boundaries.

Zhaofei and Shanshan, what is the definition of data science in your opinion? What is a data scientist? Do you each self-identify as one?

ZY: Data science is the science of data, which is defined as exploring the theories, methods, and techniques of the data world. I think it contains two meanings. First, data science is to research the data itself, including data acquisition, data storage and management, data analysis, and data visualization. Second, it mainly refers to data-driven methods, providing a new approach for natural and social sciences research. I think as long as the researcher focuses on the data itself or data-driven methods, they are a data scientist. I consider myself a data scientist and will continue working on high-dimensional neuroscience data processing.
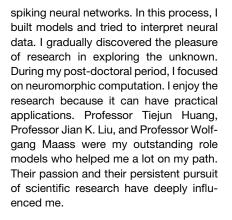
SJ: Today, data are everywhere, and data science is developing and penetrating every industry. Data science is mainly about extracting insights from data by using various technological means. In a broad sense, a researcher who mines useful information from data is a data scientist. My work focuses on various neuroscience data, including data processing and exploring the internal meaning of data, so I am confident that I am a data science researcher.

Zhaofei, what motivated you to become a (data) researcher? Is there anyone or anything in particular that helped guide you on your path?

ZY: During my PhD, my research focused on explaining how the brain implements probabilistic inference and learning with

spiking neural networks. In this process, I built models and tried to interpret neural data. I gradually discovered the pleasure of research in exploring the unknown. During my post-doctoral period, I focused on neuromorphic computation. I enjoy the research because it can have practical applications. Professor Tiejun Huang, Professor Jian K. Liu, and Professor Wolfgang Maass were my outstanding role models who helped me a lot on my path. Their passion and their persistent pursuit of scientific research have deeply influenced me.

Shanshan, what attributes make a data scientist successful?

SJ: To become a successful data scientist, you first need to have strong background knowledge of the field, rich research technology, and preferably mathematical knowledge. Specifically, researchers need to be rooted in specific research fields and must master the data characteristics in the field. In addition, with the increasing popularity of data science, new technologies are constantly updated. One should master many data-processing, data-analysis, and machine-learning algorithms. Having a mathematical foundation will help researchers better understand the data and analysis results. In addition, due to the large amount and complexity of data, it is also necessary to find clues in complex analysis results patiently and carefully. At the same time, a little luck is also essential.

Shanshan, how do you stay updated on both advances in data science techniques and the advances in the field that you work in?

SJ: At present, data science and AI can be described as the hottest fields with continuous development achievements. We update the latest advances in these fields by reading papers, following GitHub, and listening to reports, especially the new work of famous research groups and professor teams within the field. Furthermore, we often share interesting articles with members of the group through paper reading.

Zhaofei, a lot of data scientists continue their career outside of academia. What is your view on that? Do you encourage your students and postdocs to continue their career in academia and establish their own teams, and are you supportive of careers outside of academia?

ZY: I think it's a personal choice for them to choose a career outside of academia in which they can develop their strengths. Academia is still mostly about basic research, whereas the industry may be more likely to deliver results. I would encourage my students to continue their careers in academia, but I will be supportive of them if they want to pursue their career outside of

academia. It comes down to their own interests and choices.

Zhaofei, what is the fun part of being a data scientist?

ZY: I think the most interesting thing is that you have a huge amount of data. You can approach it in different ways and get different results. You find that you have many options. I really feel that "Life is like a box of chocolates. You never know what you're going to get."

REFERENCES

1. Jia, S., Li, X., Huang, T., Liu, J.K., and Yu, Z. (2022). Representing the dynamics of high-dimensional data with non-redundant wavelets. Patterns *3*, 100424.

About the authors

Zhaofei Yu is an assistant professor in the Institute for Artificial Intelligence, Peking University, Beijing, China, and he is heading the brain-inspired Vision Lab. He received his BS from the Hong Shen Honors School, College of Opto-electronic Engineering, Chongqing University, China, in 2012 and his PhD from the Automation Department, Tsinghua University, China, in 2017. After that, he completed a postdoc at Peking University (supervisor: Professor Tiejun Huang).

Shanshan Jia is a PhD student in the Institute for Artificial Intelligence, Peking University, Beijing, China. She received her BS from the College of Life Science and Engineering, Huanggang Normal University, China, in 2010 and her master's degree from the Smart City College, Beijing Union University, China, in 2018.