AI, BI & SI—Artificial, Biological and Statistical Intelligences

Dennis K.J. Lin

Distinguished Professor, Department of Statistics, Purdue University, West Lafayette, IN

ABSTRACT

Artificial Intelligence (AI) is clearly one of the hottest subjects these days. Basically, AI employs a huge number of inputs (training data), super-efficient computer power/memory, and smart algorithms to perform its intelligence. In contrast, Biological Intelligence (BI) is a natural intelligence that requires very little or even no input. This talk will first discuss the fundamental issue of input (training data) for AI. After all, not-so-informative inputs (even if they are huge) will result in a not-so-intelligent AI. Specifically, three issues will be discussed: (1) input bias, (2) data right vs. right data, and (3) sample vs. population. Finally, the importance of Statistical Intelligence (SI) will be introduced. SI is somehow in between AI and BI. It employs important sample data, solid theoretically proven statistical inference/models, and natural intelligence. In my view, AI will become more and more powerful in many senses, but it will never replace BI. After all, it is said that "The truth is stranger than fiction, because fiction must make sense." The ultimate goal of this study is to find out "how can humans use AI, BI, and SI together to do things better."

Keywords: Data Quality, Data Right and Right Data, Design of Experiment, Intelligent Data Collection Method

Solving the Mysteries of Place Cells and Grid Cells by Representation Learning

Ying Nian Wu

UCLA

ABSTRACT

The 2014 Nobel Prize in Physiology or Medicine recognized the discovery of place cells and grid cells in the mammalian brain. Each place cell fires at a single specific location, whereas each grid cell fires at multiple locations forming a hexagonal grid pattern. Yet the computational principles underlying these phenomena have remained mysterious. We show both emerge from representation learning through geometric optimization. Grid cells learn embeddings that preserve local distances through conformal isometry, forming a coordinate system. We prove hexagonal patterns are optimal: hexagonal flat tori uniquely minimize deviation from local distance preservation by distributing curvature isotropically through six-fold symmetry. Building upon this coordinate system, place cells learn embeddings that preserve spatial adjacency relations defined by transition kernels of heat diffusion with reflecting boundary conditions, thereby forming a cognitive map. Specifically, inner products between embeddings reconstruct transition probabilities, causing localized firing patterns to emerge automatically from non-negative matrix factorization constraints. This reveals how the brain solves navigation by transforming spatial reasoning into optimization on learned geometric representations.

Keywords: representation learning, computational neuroscience, spectral decomposition, non-negative matrix factorization, sparse coding.

Generate Diverse Protein Conformations through AlphaFold

Samuel Kou[1]

Department of Statistics, Harvard University

ABSTRACT

The introduction of AlphaFold has revolutionized the task of protein structure prediction from a given sequence of amino acids; the groundbreaking contribution of AlphaFold was recognized by the 2024 Nobel Prize in Chemistry. As a deep-learning based method, AlphaFold was trained from the publicly available Protein Data Bank (PDB), a database of known protein structures. An inherent limitation of AlphaFold is that its prediction can only give a static structure, whereas in reality, the structures of proteins are dynamic and can change in response to their environment or binding partners, with significant biological consequences. In this talk, we focus on enhancing and diversifying protein structure prediction using AlphaFold. Through a principled iterative statistical sampling framework, we significantly expand AlphaFold's capabilities, enabling it to explore a broader conformational space. Key methodologies involve modifying the multiple sequence alignment (MSA) and template inputs to encourage AlphaFold to explore different conformations, thereby increasing structural diversity. This is achieved in particular through an iterative sequential sampling approach, which allows for the incorporation of protein residue co-evolutionary information in the structure prediction, broadening the conformational possibilities that AlphaFold can investigate. We will illustrate the capabilities of the statistical sampling approach through examples.

Keywords: Protein folding; sequential sampling; coevolutionary information; protein conformation

Recent Advances in MM Optimization Algorithms

Hua Zhou[1], Xunjian Li[1], Kenneth Lange[2,3]

[1]*Department of Biostatistics, University of California, Los Angeles*

[2]*Department of Human Genetics, University of California, Los Angeles*

[3]*Department of Computational Medicine, University of California, Los Angeles*

ABSTRACT

The majorization-minimization (MM) principle is an extremely general framework for deriving optimization algorithms. It includes the expectation-maximization (EM) algorithm, proximal gradient algorithm, concave-convex procedure, quadratic lower bound algorithm, and proximal distance algorithm as special cases. Besides numerous applications in statistics, optimization, and imaging, the MM principle finds wide applications in large-scale machine learning problems such as matrix completion, discriminant analysis, and nonnegative matrix factorizations. This talk presents some novel applications of the MM principle in the big data setting, including parallel block least squares, deweighting weighted least squares, large-scale variance component model, independent component analysis, and multi-level Monte Carlo.

Keywords: majorization-minimization; optimization.