智源学者成果展示——人工智能的数理基础

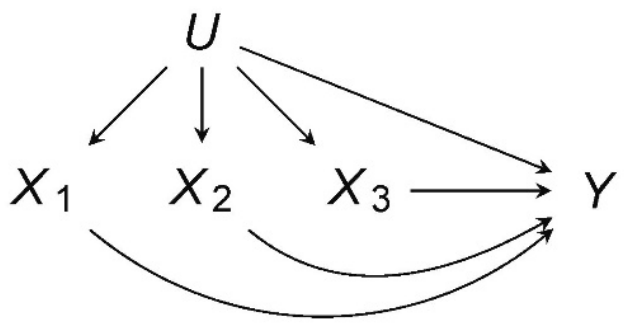
作者　林　伟（北京大学研究员、智源研究员）
朱占星（北京大学研究员、智源青年科学家）
王　涵（北京应用物理与计算数学研究所副研究员、智源青年科学家）
戴彧虹（中国科学院数学与系统科学研究院研究员、智源研究员）
邓　柯（清华大学研究员、智源研究员）

2020年6月

带隐变量的非稀疏学习

北京大学研究员、智源研究员林伟等提出一种带隐变量的非稀疏学习方法，利用因子模型和主成分分析调整稀疏学习中的混杂因素。稀疏学习因为具有良好的解释性和泛化性而被广泛使用，然而由于混杂因素的存在，个体变量效应的稀疏性假设在实际问题中往往难以满足。该论文首次从理论上证明结合因子和稀疏结构调整混杂因素是可行的，为基于观察性数据的因果推断提供了新的思路和理论框架。

Zheng, Z., Lv, J. and **Lin, W.** (2020). Nonsparse learning with latent variables. *Operations Research*, to appear.

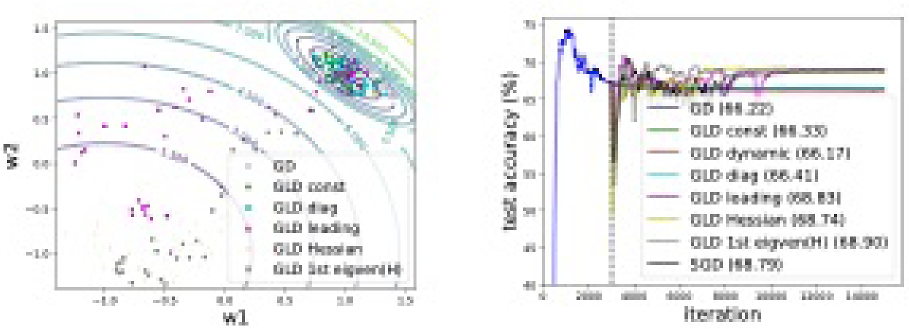


图：*U* 是 X_1, X_2, X_3 对 *Y* 效应的混杂因素，可通过因子模型来调整

深度学习训练算法的理论分析

北京大学研究员、智源青年科学家朱占星等首次发现了在深度学习中随机梯度噪声的结构对于逃离局部极小有重要作用，对于设计新的更加高效有效的训练方法有重要的指导意义。理解深度学习的训练算法的行为一直是深度学习理论的关键难题，这个发现对于理解深度学习中的随机梯度法有新的见解，部分解释了算法对于深度学习的关键作用，对于未来理解新的算法以及算法设计开启了新的可能。

Zhanxing Zhu*, Jingfeng Wu*, Bing Yu, Lei Wu and Jinwen Ma. *The Aniso-tropic Noise in Stochastic Gradient Descent: Its Behavior of Escaping from Minima and Regularization Effects*. 36th International Conference on Machine Learning (ICML 2019)

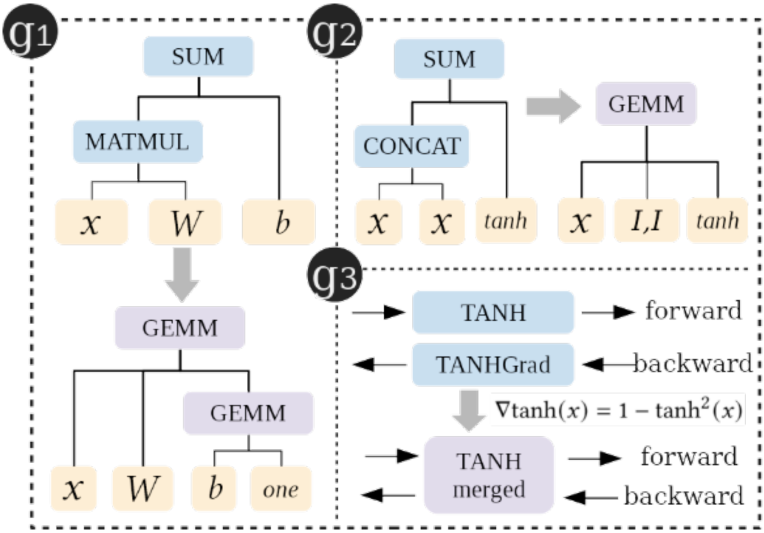


The SGD-type algorithms escape from sharp minima. Left: the trajectory of different dynamics in 2D toy example. Right: generalization performance w.r.t. iterations. 图片来自论文。

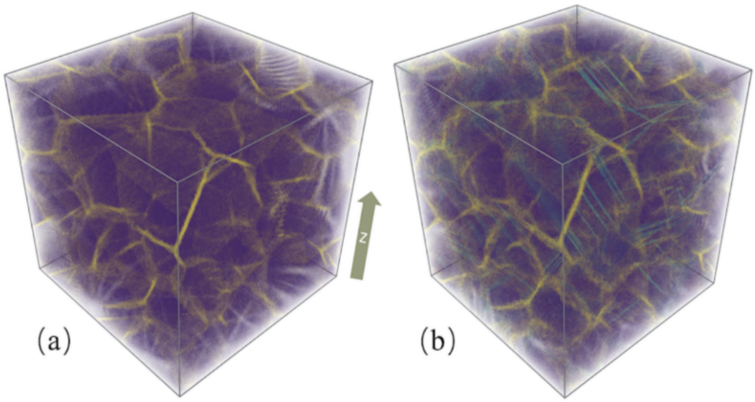
深度学习分子动力学的高性能优化

北京应用物理与计算数学研究所副研究员、智源青年科学家王涵等在 27360 张 GPU 加速卡上优化实现了深度学习分子动力学，首次实现了上亿原子体系的第一性原理精度模拟，模拟速度达到创纪录的 1 纳秒每天。这项工作相比当今最快的第一性原理精度模拟提升速度一千倍以上。

Weile Jia, Han Wang, Mohan Chen, Denghui Lu, Jiduan Liu, Lin Lin, Roberto Car, Weinan E, Linfeng Zhang, Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning, arX-iv:2005.00223, 2020



本项工作对 TensorFlow 计算图的优化。



千万原子纳米铜多晶体系的拉伸模拟。

GANs 相关的约束极小极大问题理论

中国科学院数学与系统科学研究院研究员、智源研究员戴彧虹等研究了来源于生成对抗网络 (Generative Adversarial Networks)、对抗训练 (Adversarial Training) 和多智能体强化学习 (Multi-Agent Reinforcement Learning) 中的约束极小极大问题，并在局部极小极大点 (Local Minimax Point) 的意义下给出了最优性理论。 特别地，在内层满足 Jacobian 唯一性假设下，证明了一阶和二阶必要性最优条件和二阶充分性最优条件；同时在对内层满足强正则假设下，证明了一阶必要性最优条件。

Yu-Hong Dai and Liwei Zhang, Optimality Conditions for Constrained Minimax Optimization, arXiv:2004.09730v1 (Accepted by CSAM), 2020.

Theorem 3.1 *(Necessary Optimality Conditions) Let $(x^*, y^*) \in \Re^n \times \Re^m$ be a point around which f, h, g are twice continuously differentiable and H, G are twice continuously differentiable around x^*. Let (x^*, y^*) be a local minimax point of Problem (1.1). Assume that the linear independence constraint qualification holds at y^* for constraint set $Y(x^*)$. Then there exists a unique vector $(\mu^*, \lambda^*) \in \Re^{m_1} \times \Re^{m_2}$ such that*

$$\begin{aligned} &\nabla_y \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) = 0, \\ &h(x^*, y^*) = 0, \\ &0 \geq \lambda^* \perp g(x^*, y^*) \leq 0. \end{aligned} \tag{3.4}$$

For any $d_y \in C_{x^*}(y^*)$, we have that

$$\langle \nabla^2_{yy} \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) d_y, d_y \rangle \leq 0. \tag{3.5}$$

Assuming Problem (P_{x^*}) satisfies Jacobian uniqueness conditions at (y^*, μ^*, λ^*) and the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ, there exists $(u^*, v^*) \in \Re^{n_1} \times \Re^{n_2}$ such that

$$\begin{aligned} &\nabla_x \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) + \mathcal{J} H(x^*)^T u^* + \mathcal{J} G(x^*)^T v^* = 0, \\ &H(x^*) = 0, \\ &0 \leq v^* \perp G(x^*) \leq 0. \end{aligned} \tag{3.6}$$

The set of all (u^*, v^*) satisfying (3.6), denoted by $\Lambda(x^*)$, is nonempty compact convex set. Furthermore, for every $d_x \in C(x^*)$ where $C(x^*)$ is defined by (3.3),

$$\begin{aligned} \max_{(u,v)\in\Lambda(x^*)} &\left\{ \left\langle \left[\sum_{j=1}^{n_1} u_i \nabla^2_{xx} H_j(x^*) + \sum_{i=1}^{n_2} v_i \nabla^2_{xx} G_i(x^*) \right] d_x, d_x \right\rangle \right\} \\ &+ \left\langle \left[\nabla^2_{xx} \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) - N(x^*)^T K(x^*)^{-1} N(x^*) \right] d_x, d_x \right\rangle \geq 0, \end{aligned} \tag{3.7}$$

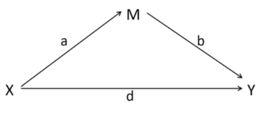
where $K(x)$ is defined by (2.6) and $N(x)$ is defined by

$$N(x) = \begin{bmatrix} \nabla^2_{x,y} \mathcal{L}(x; y(x)\mu(x), \lambda(x)) \\ 0 \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix}. \tag{3.8}$$

中介效应因果推断不依赖于总和因果效应检验

清华大学研究员、智源研究员邓柯等首次严格证明了在很宽泛的条件下，“总和因果效应检验”（Total-Effect Test）对于建立中介因果效应是不必要的。通过深入分析中介效应模型中一系列参数检验所对应的检验拒绝域之间的几何关系，将总和因果效应检验的必要性问题转化为对相关检验拒绝域的几何分析问题，成功加以解决。该成果对于理解在复杂系统下的因果机制具有重要的理论和现实意义，为理清了困扰学界长达 30 年之久的学术争论提供了重要的理论依据。

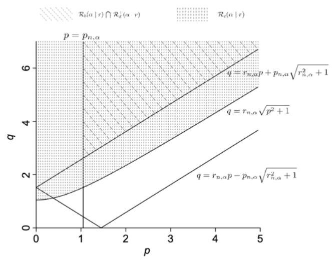
Yingkai Jiang, Xinshu Zhao, Lixing Zhu, Jun S. Liu & **Ke Deng**. *Total-effect test is superfluous for establishing complementary mediation*. To appear in *Statistica Sinica*



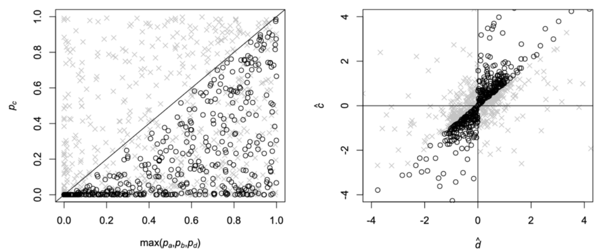
(a) 中介效应因果机制

Test	Model	Hypotheses	Rejection region of F-test
$\mathcal{T}_a$	$M = i_M + aX + \varepsilon_M$	$H_0: a = 0, H_1: a \neq 0$	$\mathcal{R}_a(\alpha) = \left\{ \frac{\lVert\mathbf{M}_{1,\mathbf{X}} - \mathbf{M}_1\rVert/(2-1)}{\lVert\mathbf{M} - \mathbf{M}_{1,\mathbf{X}}\rVert/(n-2)} > \lambda_{1,n-2}(\alpha) \right\}$
$\mathcal{T}_b$	$Y = i_Y + bM + dX + \varepsilon_Y$	$H_0: b = 0, H_1: b \neq 0$	$\mathcal{R}_b(\alpha) = \left\{ \frac{\lVert\mathbf{Y}_{1,\mathbf{M},\mathbf{X}} - \mathbf{Y}_{1,\mathbf{X}}\rVert/(3-2)}{\lVert\mathbf{Y} - \mathbf{Y}_{1,\mathbf{M},\mathbf{X}}\rVert/(n-3)} > \lambda_{1,n-3}(\alpha) \right\}$
$\mathcal{T}_d$	$Y = i_Y + bM + dX + \varepsilon_Y$	$H_0: d = 0, H_1: d \neq 0$	$\mathcal{R}_d(\alpha) = \left\{ \frac{\lVert\mathbf{Y}_{1,\mathbf{M},\mathbf{X}} - \mathbf{Y}_{1,\mathbf{M}}\rVert/(3-2)}{\lVert\mathbf{Y} - \mathbf{Y}_{1,\mathbf{M},\mathbf{X}}\rVert/(n-3)} > \lambda_{1,n-3}(\alpha) \right\}$
$\mathcal{T}_c$	$Y = i_Y^* + cX + \varepsilon_Y^*$	$H_0: c = 0, H_1: c \neq 0$	$\mathcal{R}_c(\alpha) = \left\{ \frac{\lVert\mathbf{Y}_{1,\mathbf{X}} - \mathbf{Y}_1\rVert/(2-1)}{\lVert\mathbf{Y} - \mathbf{Y}_{1,\mathbf{X}}\rVert/(n-2)} > \lambda_{1,n-2}(\alpha) \right\}$

(b) 与建立中介效应相关的统计检验



(c) 对统计检验拒绝域的几何关系进行分析



(d) 理论证明的结论得到了数值模拟的验证

运用几何分析方法梳理“总和因果效应检验”（Total-Effect Test）对于建立中介因果效应的必要性。
图片来源：作者论文

Beijing Academy of
Artificial Intelligence

微信关注
北京智源人工智能研究院