

第五届全国统计物理和复杂系统学术会议 安徽 合肥
中国科学技术大学

机器学习方法研究渗流和XY模型
Phys. Rev. E 99, 032142 (2019)

张万舟副教授

太原理工大学物理与光电工程学院

2019年7月26日

合作者:

Tzu-Chieh Wei 教授(纽约州立大学石溪分校, 杨振宁理论物理研究所)

刘嘉钰(硕士生, 已毕业)

一、研究问题

- 1、机器学习识别非局域特征和拓扑相变
- 2、有没有其他的输入方案，避免神经网络的重复训练？

二、研究方法

- 1、无监督方法(t-Distributed Stochastic Neighbor Embedding, t-SNE)
- 2、全连接神经网络(fully-connected network, FCN)
- 3、卷积神经网络(convolutional neural network ,CNN)
- 4、直方图输入方案

三、数值结果和结论

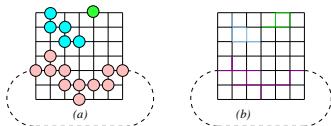
点渗流和棒渗流、XY模型和广义XY模型($q=2, 3$ 和8)

一、研究问题

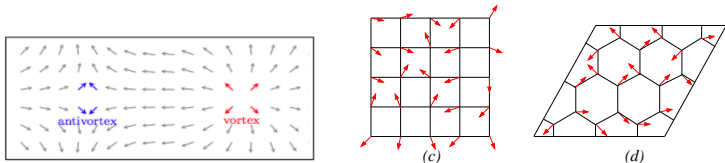
J. Carrasquilla and R. Melko, Nat. Phys. (2017) 机器学习伊辛模型, 其中 $Outputlabel = f(Input\{S_i\})$ 与 $m = |\sum S_i/N|$ 类似具有局域特征

问题1: 机器学习能否识别具有非局域特征或者拓扑特征的相、并预测相变点? 以渗流和XY模型为例

$$Z = \sum_{\{\sigma\}} p_s^{n_s^\sigma} (1 - p_s)^{N - n_s^\sigma}$$



点(棒)渗流: 连通或不连通具有非局域特征



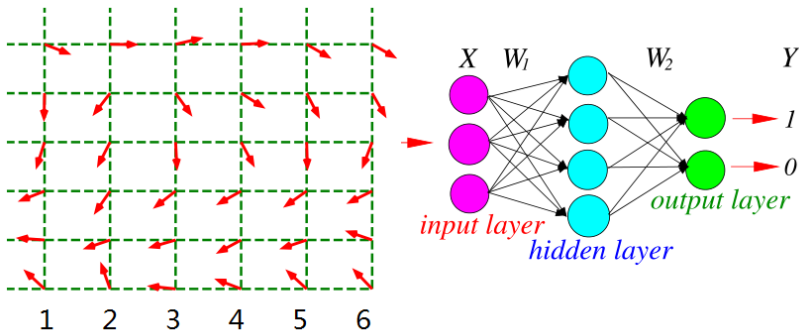
XY模型拓扑相变: 正(反)涡旋 正方晶格 六角晶格

一、研究问题

神经网络的输入:

神经网络输入层的神经元个数和输入图像大小相等。在识别物质的相时，需要计算不同尺寸的系统。不同尺寸的系统又需要**重新**训练网络。尺寸越大，越耗时间和内存。

因此，除了优化网络结构的方案外，**问题2：对于不同的尺寸，如何避免重新训练网络？**



二、研究方法

t-SNE方法 降维机器学习算法

$N=4 \times 4$ 正方晶格点渗流, 10个概率, 每个概率生成1000个样本

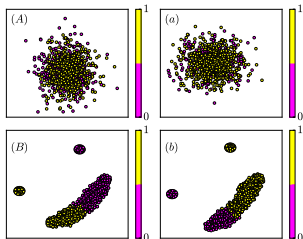
概率	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
标签	0	0	0	0	0	1	1	1	1	1

输入矩阵为 $10,000 \times 16 \Rightarrow$ 输出 $10,000 \times 2$:

```

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1
1 0 1 1 0 1 0 0 0 1 1 0 0 0 0 0
1 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0
0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0
1 1 0 1 1 1 0 0 1 1 1 1 0 1 1
1 1 1 1 0 1 1 1 1 1 1 1 0 1 1
0 1 1 1 1 1 0 0 1 1 1 1 1 1 1
1 1 1 1 0 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```



两种颜色分别表示连通和不连通

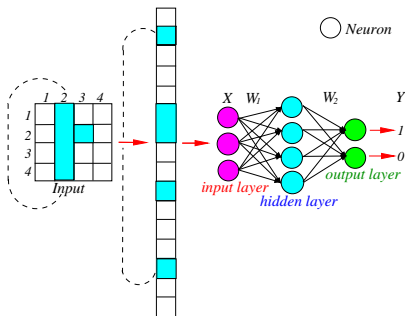
结论: t-SNE 方法可以成功对渗流数据进行降维, 清楚的看到渗流、非渗流和临界区域。

FCN方法

全连接神经网络(fully-connected network, **FCN**)

目的：建立输入 x 输出 y 之间的关系 $y = g(f(x))$

$x \Rightarrow y_H \Rightarrow y$: $y_H = f(x \cdot W_1 + b_1)$, $y = g(y_H \cdot W_2 + b_2)$



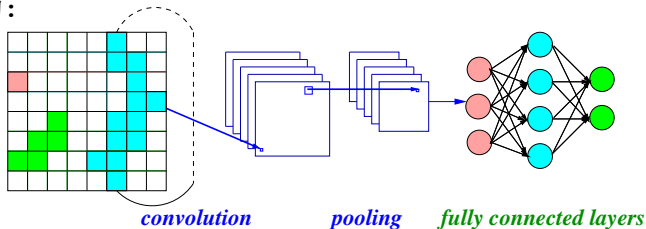
训练阶段: 通过**已知**数据集 x, y 训练网络, 得到连接权重 W 和偏离量 b , 建立函数映射关系 f, g ; **测试阶段**: 输入某个渗流构型, 网络判断**连通的概率为1**, **不连通的概率为0**, 和预期一致。

CNN方法

卷积神经网络(convolutional neural network, CNN)

卷积层+池化层的个数, 可根据实际效果调节。

网络结构:



卷积细节

INPUT IMAGE					
18	54	51	239	244	188
55	121	75	78	95	88
35	24	204	113	109	221
3	154	104	235	25	130
15	253	225	159	78	233
68	85	180	214	245	0

WEIGHT

1	0	1
0	1	0
1	0	1

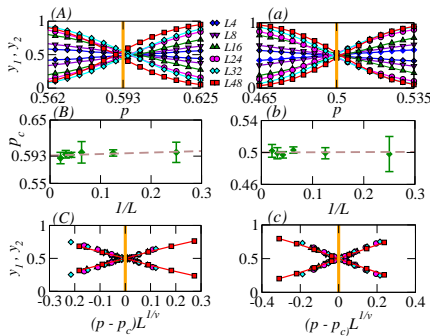
429

$$18 \times 1 + 51 \times 1 + 171 \times 1 + 35 \times 1 \\ + 204 \times 1 = 429$$

三、数值结果和结论

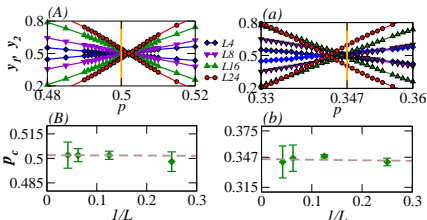
FCN结果

点渗流和棒渗流，正方(三角)晶格，相变点，临界指数，数据跌落， $p \Rightarrow (p - p_c)L^{1/\nu}$ ，其中 $\nu = 4/3$ 。 y_1, y_2 是输出层的平均值，处于两个相(渗流和非渗流)的概率。相变点处， $y_1=y_2=0.5$ ；
 $p = 0(1)$ 时 $y_1 = 0(1)$, $y_2 = 1(0)$ ；



正方晶格点渗流

三角晶格

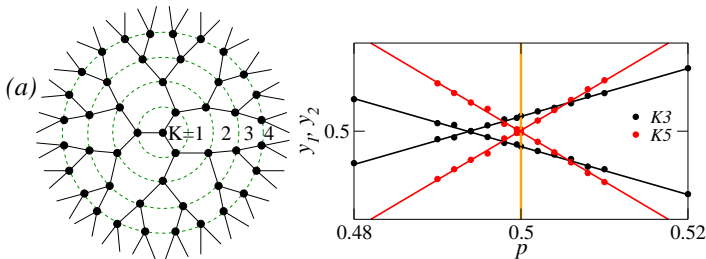


正方晶格棒渗流

三角晶格

FCN结果

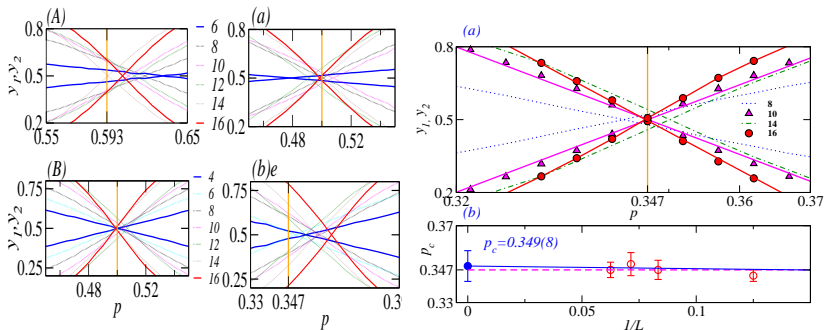
非常规晶格 Bethe晶格点渗流，严格的相变点为 $p_c = 0.5$ ；与正
方（三角）晶格不同的是，Bethe晶格具有拓扑树结构，从中心
点向外拓展。判断任意两个不同分支上的点是否连通，也具有非
局域特征。



A. Saberi, Recent advances in percolation theory and its applications,
Phys. Rep. (2015).

CNN结果

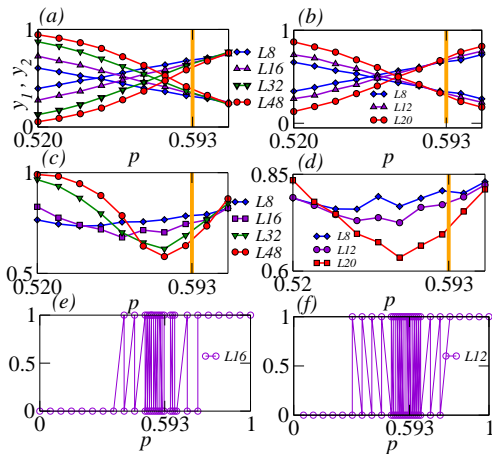
图(b)效果不佳：按照均匀分布的占据概率 p 生成的构型去训练网络，比如 $p = 0.1, 0.2, 0.3, 0.4, 0.5, \dots$ 怎么改进结果？



(A) (a) 正方晶格和三角晶格点渗流 (B) (b) 三角晶格棒渗流

改进方案：按照以 $p_c = 0.347$ 为对称中心,但是不包括 p_c 的概率列表生成构型去训练网络, $\dots, p_c - 3\delta, p_c - \delta, p_c + \delta, p_c + 3\delta, \dots$ 结果为 $p_c = 0.349(8)$, 和精确值0.347一致。

尝试真实标签： 之前的规则，参数 p 下的构型 \mathbf{x} ，在 $p > p_c$ 时，标签 $\mathbf{y}=\mathbf{1}$ ； $p < p_c$ 时 $\mathbf{y}=\mathbf{0}$ 。然而，真实的标签在相变点附近是存在涨落的，如(e) 和(f) 所示。



真实的标签，即使存在涨落，但是仍然可以较好的训练网络。

激活函数是否会引起系统误差？

sigmoid $\sigma(z) \equiv \frac{1}{1+e^{-z}}$ 与费米狄拉克分布有关 $z = -(\epsilon - \epsilon_f)/K_B T$

softmax $\mathbf{a}_j = \frac{e^{z_j}}{\sum_k e^{z_k}}$ 与玻尔兹曼分布有关 $z = -\epsilon/K_B T$

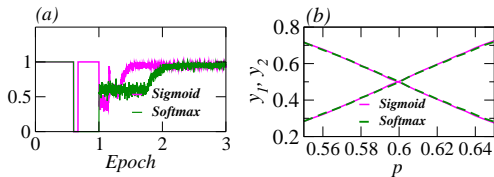


Figure: (a) The training accuracy by using the sigmoid and softmax activation, respectively. (b) For the site percolation on the square lattices, we find the average output layers are the same using enough samples (25000).

对于渗流模型，我们没有发现系统误差。

结论1

运用机器学习方法研究了二维晶格的点渗流和棒渗流模型。

对于渗流相变，非监督t-SNE 方法可以把一系列构型形成的高维数据集映射到二维可分类的数据图像。

利用FCN,不用考虑空间结构，渗流相变仍然可以被识别。

通过输入构型的真实标签，即构型否存在渗流集团，而不用根据生成的构型时 p 是否大于或者小于 p_c ，也可以较好的训练网络。

通过输出神经元的期望值的数据跌落，我们可以得到正确的临界指数 ν 。

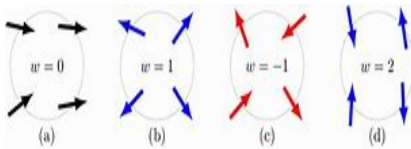
XY 模型

经典XY模型: $H = -J \sum_{\langle i,j \rangle} \vec{s}_i \cdot \vec{s}_j = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j)$,
低温时涡旋成对出现, 高温时涡旋对解散, 拓扑相变。

Hu, Singh, and Scalettar, PRE. (2017) “We believe, seeing the vortex- Antivortex unbinding transition in 2D XY model **may be difficult in PCA or any machine learning technique.**”

Beach, Golubeva, Melko, Machine learning vortices, PRB (2018).
用涡旋作为输入, 要1000个自旋以上, 精度才可以接受。

Suchsland, Wessel, Parameter diagnostics, PRB (2018).



零涡旋 涡旋+1 反涡旋-1

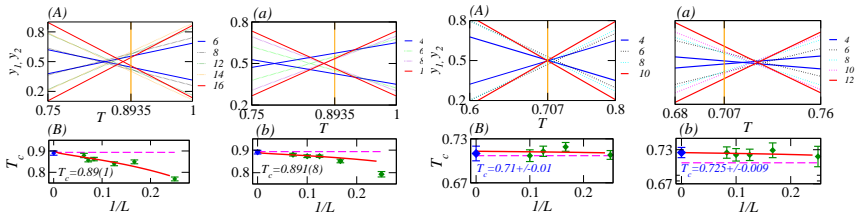
Roger Wessel

XY 模型

输入:

正方 $\{\theta_i\}$, $\{\sin \theta_i, \cos \theta_i\}$

六角 $\{\theta_i\}$, $\{\sin \theta_i, \cos \theta_i\}$?

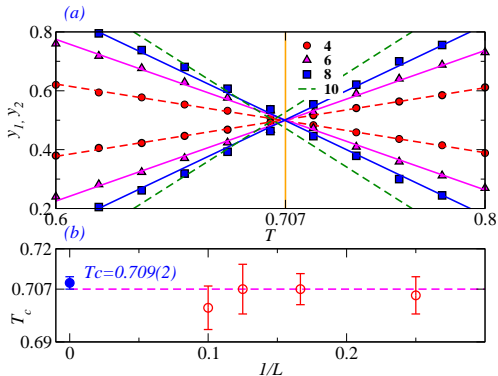


$$T_c(L) = T_c + \frac{b}{(\log(L))^2}, \quad \text{Nelson 和 Kosterlitz, PRL (1977).}$$

神经网络的performance在相变点附近, 准确度相对较低。为了减小标准偏差, 每个温度用了25000 训练样本和测试样本。这里的测试样本, 不是Validation样本。

改进方案：按照以 $T_c = 0.707$ 为对称中心,但是不包括 T_c 的**温度列表**生成构型去训练网络

温度列表: ..., $T_c - 3\delta$, $T_c - \delta$, $T_c + \delta$, $T_c + 3\delta$, ...



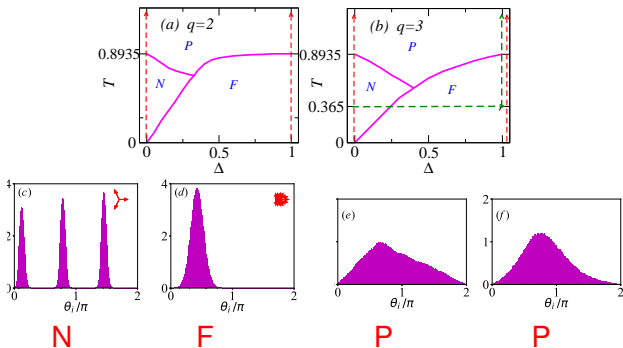
$T_c = 0.709(2)$ 和精确值0.707 一致。

广义XY (GXY) 模型

$$H = - \sum_{\langle i,j \rangle} [\Delta \cos(\theta_i - \theta_j) + (1 - \Delta) \cos(q\theta_i - q\theta_j)],$$

权重比例 $\Delta = 1$ 时, GXY 模型 \rightarrow XY 模型。

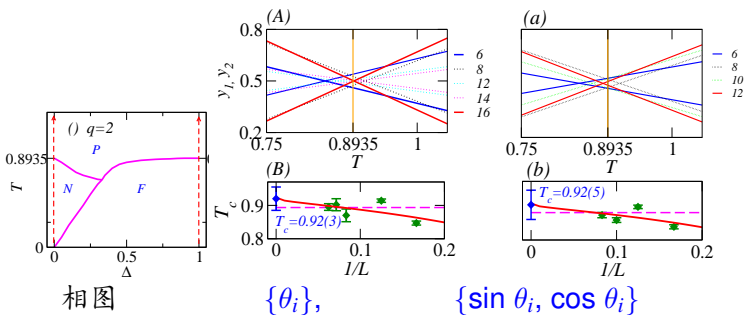
相图以及自旋角度分布:



N nematic (N)相, 向列相, 即自旋角度的定向排列。 $q=3$, 自旋角度的统计分布有3个尖峰。**F**: 准长程序铁磁相。**P**: 顺磁相。

广义XY (GXY) 模型

$q = 2$, $\Delta = 0$ 扫描温度 T :

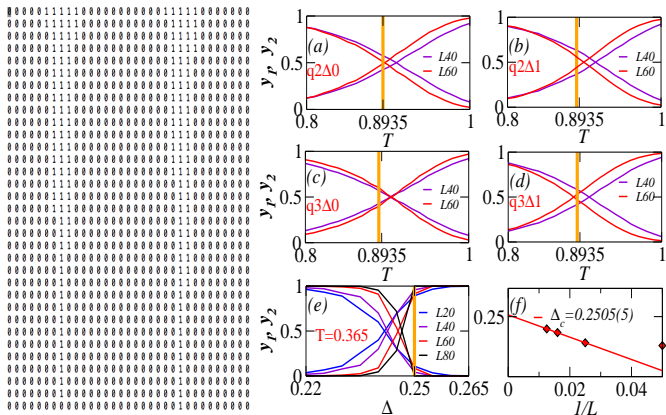


在 $\Delta = 0$ 时, 令 $q\theta_i \rightarrow \bar{\theta}_i$, $\sum_{i,j} -\cos(q\theta_i - q\theta_j)$ 和 XY 模型同构, 具有相同的 T_c 。Y. Levin et al, arXiv:1207.3447.

模拟的系统尺寸能否加大? 然而, 32×32 , 25000 训练样本已经足够慢, 甚至只能算 1-2 个任务。

广义XY (GXY) 模型

统计分布图作为输入: 把角度的统计直方图进行剪切, 然后切割为 32×32 图像, 分布线内部用1表示, 外部用0表示.



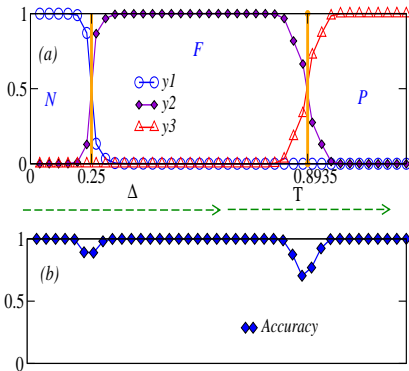
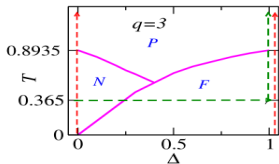
好处1; 可以把系统 $L = 20, 40, 60, 80$ 甚至 $150, 200$ 的自旋统计直方图切割为 32×32 , 节省计算资源

广义XY (GXY) 模型

推广到3个输出神经元，运用 softmax 函数， $z_j \Rightarrow y_j = \frac{e^{z_j}}{\sum_k e^{z_k}}$

先沿着温度 $T = 0.365$ 扫描，接着沿着 $\Delta = 1$ 扫描；为了让网络感受到明显的特征，训练网络时，避开三临界区域。

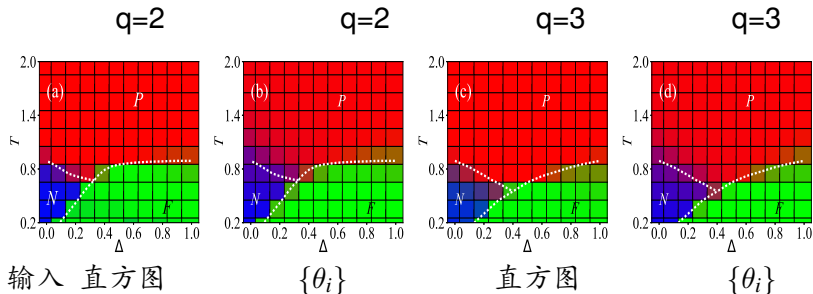
相	N	F	P
y_1	$\neq 0$	0	0
y_2	0	$\neq 0$	0
y_3	0	0	$\neq 0$



能否给出全局相图？

广义XY (GXY) 模型

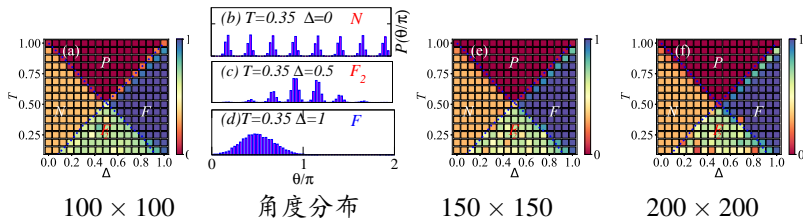
q=2,3 全局相图 运用相图中的部分数据，沿着 $T = 0.365$ 以及 $\Delta = 1$ 得到的样本，训练神经网络。测试时，输入**全局相图** ($0 < \Delta < 1$ 且 $0.2 < T < 2$) 的样本，输出其标签。 y_1, y_2, y_3 恰好代表蓝、绿、红色的权重。



虚线数据来自 Hübcher and Wessel, PRE(2013); Poderoso, Arenzon and Levin, New ordered phases in a class of generalized XY models, PRL (2011).

q=8 全局相图

Canova, Levin, and Aizenberg, Competing nematic interactions in a generalized XY model ..., PRE (2016); Kosterlitz-Thouless and Potts transitions in a generalized XY model, PRE(2014). 表明 $q > 3$ 系统存在新的相 F_2



好处2: 训练尺寸为 100×100 样本的网络，给出尺寸为 150×150 , 200×200 的相图，而且和已知相图完全一致，避免重复训练网络，每个尺寸的系统节约1周时间。

结论2

正方和六角晶格的XY模型，输入 $\{\cos\theta_i, \sin\theta_i\}, \{\theta_i\}$ ；即使尺寸很小 $L = 4, \dots, 16$ ，通过有限尺寸标度，仍然可以得到临界点。 $q = 2, 3, 8$ 的广义XY模型，训练网络时，用了一小部分极限参数下的数据，输出整个相图。 $q = 8$ 时，在 $0 < \Delta < 1$ 范围内，从机器学习角度，验证新的相 F_2 。

自旋构型可以作为神经网络的输入，然而不同相之间的最大差别是自旋角度的直方图分布。因此，我们设计了基于一种自旋直方图输入方案，成功实现XY模型和广义XY模型的相变和相图预测。

此外，运用尺寸为 $L \times L$ 样本的训练网络，测试其它尺寸($L' \times L'$, $L' \neq L$) 的数据，节约额外的训练成本和时间。

The use of machine learning in phases of matter in general may still need the ingenuity of appropriate features for the neural network to learn, but can become a useful tool.

感谢您的关注