

机器学习在座逾渗相变问题中的应用

徐荣幸,赵鸿

(厦门大学 物理系,福建 厦门 361005)

摘要:机器学习理论区别于传统方法,因其在对于复杂的数据集识别、分类的准确性和高效性而被广泛应用于各个领域.识别相变是机器学习和统计物理领域相结合的最有代表性的工作.到目前为止,机器学习完成的相变识别几乎都是基于具有动力学演化过程的自旋模型,如 Ising 模型等,而其在另一类不具有动力学演化过程而完全由系统结构特征决定的相变模型,如逾渗模型等,仍未有细致研究.本文结合现有的机器学习技术,卷积神经网络和一般向量机,对二维方格子上的座逾渗问题进行了研究,发现能以高正确率对不同相的构型进行识别,证明了机器学习在这类问题上研究的可行性.通过已完成训练的学习机对不同参数下构型预测的正确率计算,发现正确率在相变点附近会出现急剧衰减,与系统参数呈幂律衰减.这与传统相变理论一致.通过定量计算,还发现 2 种学习机的正确率衰减规律都满足同一个幂律指数.这不仅进一步从全新的角度揭示了相变的普适性,而且为找寻相变点提供了新的方法.

关键词:机器学习;卷积神经网络;一般向量机;座逾渗;相变;正确率;标度关系

中图分类号:O414.21;TP319.4

文献标志码:A

近年来,机器学习特别是深度学习飞速发展,给几乎所有的领域带来了一场巨大的变革.比如物体识别,医疗诊断和自然语言处理等等.这些都归因于它在复杂的数据集上表现良好的识别,判断以及分类能力.特别是其在预测晶体结构^[1],量子多体问题^[2]以及分类相变^[3-8]上取得了成功,使得机器学习逐渐成为了解决物理问题的崭新工具.

在这个结合领域中,发展最为迅速和成熟的是机器学习和相变问题的结合.到目前为止,这类问题可按机器学习的类型分为 2 类.1)无监督学习方法.Wang 等人绕开复杂的物理理论,直接从数据出发,利用主成分分析的方法,发现 Ising 模型的高温相和低温相可以直接被分开,并且发现降维投影的过程正是同求解 Ising 模型的序参量过程一致^[7].van Nieuwenburg 等人提出了专门针对相变问题的无监督学习方法,并且证明在自旋模型上,比如 Ising 模型、Kitaev 模型上该方法都能有效地寻找到相变点^[5].这两者的做法开创了只用数据研究物理而绕开物理理论的先河.这是此前物理学领域本身没有出现过的思路.2)利用监督学习方法.典型的工作是 Carrasquilla 和 Melko 利用单隐含层神经网络和卷积神经网络对 Ising 模型、自旋冰模型和 Ising 规范模型进行的监督学习方法,体现了神经网络强大的特征提取能力^[3].同样在 2017 年,被 *Physical Review Letters* 杂志评为 2017 年十大进展的工作也是关于相变问题的.他们提出了一种结合量子图方法的预处理数据方法,为机器学习应用到一类特定的拓扑相变上的失败训练提供新的解决方案^[8].机器学习在拓扑相变模型和经典 Ising 相变模型上的成功,为一些难以定出相变点的问题提供了新的解决方案^[5].

但到目前为止,现有的机器学习对相变问题的研究几乎都是基于自旋模型上的.这类自旋模型的共同特征是,它们都由其各自的哈密顿量决定其性质,需通过一定的方式演化到达系统的平衡态后,在不同的参数下才会出现不同的相.也就是说,这类模型在演化过程中,它们的不同相会根据演化方式逐渐在特征空间中发

收稿日期:2018-11-12;修回日期:2018-11-23.

基金项目:国家自然科学基金重点项目(11335006)

作者简介:徐荣幸(1993-),男,四川达州人,厦门大学硕士研究生,研究方向为统计物理,E-mail:xurongxing@stu.xmu.edu.cn.

通信作者:赵鸿,E-mail:zhaoh@xmu.edu.cn.

生自动聚类,使得不同的相对于学习而言变得比较容易区分.比如在 Ising 模型中,系统的序参量被定义为该系统的平均磁化率 $\langle M \rangle$,即单位格点上的平均自旋大小: $\langle M \rangle = \frac{1}{N} \sum_{i=1}^N s_i$.在不同温度下,系统经过动力学演化,将自发地达到 2 种不同的状态中的一种,即所有自旋方向一致(铁磁态)和上下自旋数目接近(顺磁态).对于学习机来说,仅仅只需要学得如何统计上下自旋的数目即可轻易区分这 2 种相.但是另一类相变模型,如逾渗相变,它们并没有动力学演化过程,而是随机结构上的自发产生的不同的相.与自旋类模型相比,这类模型的相的定义有一些主观成分,因此不同的相之间的区分度不如自旋类的经过动力学演化的相那么明显.对学习机而言,区分难度相对较大.最近也有使用不同于本文的神经网络来对这类问题进行研究的工作,但未强调自旋模型与逾渗模型在动力学性质上的本质差异^[6].在这样的背景下,学习机究竟能否通过训练来正确区分更为复杂的相,还需作更深一步的研究.

为回答上述问题,本文利用近年在图像处理上表现优异的卷积神经网络(Convolutional Neural Network, CNN)和一种基于 Monte-Carlo 方法优化的单隐含层神经网络——一般向量机(General Vector Machine, GVM),对二维方格子上的座逾渗相变问题做了细致的研究.首先用这 2 种网络作为监督学习机,以二维方格子的座逾渗构型作为训练和测试集,成功训练学习机能够正确识别逾渗相和非逾渗相.以此为基础,还重现了逾渗模型的相变曲线.它能与传统方法模拟的相变曲线基本重合.更进一步,通过学习机对某些参数点下的相的识别正确率的分析,发现正确率曲线会在相变点参数附近发生衰减,并且衰减随参数变化服从幂律关系.基于传统相变理论的思想,定量地计算了这个幂律指数,发现其在 2 种网络中是几乎相等的.这不仅验证了相变点的临界行为,进一步揭示了相变的普适性,而且还说明了学习机的正确率可以被用来作为新的参量研究相变,为相变研究提供了全新的思路.

1 二维方格子上的座逾渗问题

逾渗模型是由 Broadbent 和 Hammersley 最早为解决流体通过多孔介质而提出的^[9-11].在一个二维方格子 Z^2 上,每个格子都独立地以概率 p 打开, $1-p$ 关闭.问题是,当概率 p 到达什么值的时候,系统存在一个开放通路,使得在这条通路上的所有格子存在从一条边界连接到对边的情况.这种模型在逾渗问题中被称之为座逾渗问题.而另一类更为常见的模型,也是定义在格点系统上.不同于前述在格子上的独立开关,该模型是定义格点与格点之间的连接键是否以概率 p 开关,最后问题也为是否存在一条从一条边界连接到对边的格点与格点相连的大集团.这种模型在逾渗问题中被称为键逾渗问题.已有数学证明,所有的键逾渗结构都可以通过某种方式等价于座逾渗结构,而存在一些座逾渗结构却无法转化为键逾渗结构^[13].

不论是键逾渗问题还是座逾渗问题,其临界现象都是这类问题最明显的标志,即当系统尺寸无限大时,存在一个阈值 p_c , $p > p_c$ 时,这样的通路才能被观察到.绝大多数逾渗问题可以通过上述的键座逾渗的相互转换和对偶方式对问题进行简化,然后得出解析解,然而到目前为止二维方格子上的座逾渗问题在严格解析上还没有被解决.逾渗模型的这种临界现象是非线性物理学中的难题之一.在实际的应用中,逾渗模型被成功地用于无序系统上的输运问题,比如森林大火的蔓延^[12],随机介质中的输运^[10],随机网络中电路传导^[14]等.逾渗模型的物理性质丰富,是一类重要的相变问题.

除了解析方法,人们还使用 Monte-Carlo 方法来数值计算逾渗的阈值.在二维方格子上的座逾渗问题中,数值计算的相变点的结果大约为 0.592 7^[15].通常描述逾渗临界现象的物理量主要有逾渗概率 P_∞ 、平均集团大小 S 和关联长度 ξ ^[13].

定义逾渗概率 P_∞ 为构型中逾渗集团大小 N_∞ 占总开放格子 N_{open} 的比例

$$P_\infty = N_\infty / N_{\text{open}}.$$

接着是平均集团大小.假设有 N_s 个大小为 s 的集团存在于总格点数为 N 的构型,那么该构型中的集团大小分布 n_s 为 N_s/N .在这里,统计集团大小时都没有将逾渗集团包含在内.因此,可以求得一个开放格点正好处于大小为 s 的集团的概率 $w_s = \frac{N_s n_s}{N \sum_s n_s} = \frac{s n_s}{\sum_s s n_s}$.故平均集团大小

$$w_s = \frac{N_s n_s}{N \sum_s n_s} = \frac{s n_s}{\sum_s s n_s}.$$

$$S = \sum_s s w_s = \frac{\sum_s s^2 n_s}{\sum_s s n_s}.$$

定义大小为 s 的集团中第 i 个格点的位置 r_i ,可以得到格点的平均距离 $\bar{r} = \frac{\sum_{i=1}^s r_i}{s}$. 由此该集团的半径

$$R_s^2 = \frac{1}{s} \sum_{i=1}^s (r_i - \bar{r})^2.$$

所以关联长度 ξ 可由下式得到:

$$\xi^2 = \frac{\sum_s s^2 n_s \langle R_s^2 \rangle}{\sum_s s^2 n_s}.$$

这 3 个参量中,关联长度在相变理论的有限尺寸标度方法中起决定性作用.根据临界点附近 3 个参量的幂律逼近行为,可以定义 3 者分别对应的临界指数 β, γ 和 ν .

$$P_\infty \sim (p - p_c)^\beta, (p > p_c), S \sim |p - p_c|^{-\gamma}, \xi \sim |p - p_c|^{-\nu}.$$

这些临界指数仅与格子的维度和拓扑结构有关,与是否是键逾渗还是座逾渗无关.因此二维方格子座逾渗问题和键逾渗问题遵循同样一组临界指数,这表明它们属于同一种普适类.这组指数为:

$$\beta = 5/36, \gamma = 43/18, \nu = 4/3.$$

而在机器学习研究中,无法通过机器学习给出上述的几个物理量.因为在分类问题中机器学习仅仅能对已产生的构型进行分类.故,在接下来的研究中,是定义在某参数点 p 下,出现逾渗集团的构型数 n_1 占在该点产生的所有构型数 n 的比例 $P(p)$ 作为研究的物理量.其定义为:

$$P(p) = \frac{n_1}{n}.$$

每个构型都是在给定一个打开概率下随机地生成的.这些打开概率均匀分布在取值范围 $[0, 1]$ 内.产生一个构型时,构型上的每个格点都使用同一个 p , 并且每个格点都是相互独立地被设置为打开或者关闭.打开的格子将被记作“1”,而关闭的格子会记作“0”,这样每一张构型都是一个只含有 0 和 1 的矩阵.使用 Two-pass 连通域划分算法^[16]在每个构型中确认其是否逾渗,用 0 表示“逾渗”,用 1 表示“不逾渗”.

2 卷积神经网络和一般向量机

本文使用到的学习机为卷积神经网络和一般向量机.下面将分别简述 2 种学习机的结构和功能.卷积神经网络是专门用于处理和识别图像的人工神经网络.区别于其他神经网络,隐含层中含有其特有的卷积层和池化层.卷积层即是对输入作卷积操作的层,通过采用局部感受野的方法,可限制隐含层神经元和输入的连接.这种局部连接的方式构成过滤器.再将过滤器与对应的图像像素区域作权值相加(即卷积操作),可以提取这一小块区域在这一过滤器下的某一个特征.池化层与卷积层的工作原理类似物理学上的粗粒化过程,将图像的临近像素点,通过求过滤器内的平均值或者最大值等方式合并成一个点.这个结构是在处理大型图片时为加快运算速度而设计的.另外,卷积神经网络还需引入权值共享.即,在假设图像的某个局部统计特性都相同的前提下,将每一个过滤器的参数从局部共享到整张图像上.在这种情况下,不同的过滤器可以看作不同的图像特征提取方式.在同一个卷积层中,采用不同数量和大小的过滤器,就可以提取到图像不同层面的信息.将这些特征综合起来并进一步处理,即可完成识别.局部感受野和权值共享 2 种假设使得卷积神经网络摆脱了冗杂的网络参数,使得计算速度进一步加快.并且局部感受野对于局部信息的提取和关联信息的保护,使得学习机的准确性也有了极大提高^[17].

本文采用的卷积神经网络如图 1 所示.输入层是一系列边长为 L 的二值图像,因而实际输入过程中是 $L \times L$ 的矩阵.卷积层包含 2 个部分.上半部分使用了 8 个 3×3 的卷积核来对输入的局部特征进行提取,因此对于每一个构型一共可以得到 8 个不同的局域特征.下半部分采用了 Inception 的思想.将 2 个拥有 8 个 3×3 卷积核的小卷积层叠在一起,看作与上半部分同层的小结构.这种 Network-in-Network 的思想^[18]可以

加深网络隐含层数,使网络具有更强的非线性和特征提取能力.2 个连续使用的 3×3 的卷积核可以看作直接使用 5×5 的卷积核.但是前者相比后者,参数量有所减少,因此可以加快网络的训练速度.将所得的 16 个卷积层处理的特征图像整合到一起.下一层是具有 256 个神经元的全连接神经网络,主要是对 16 张特征图片的信息进行综合整理.于是输入的构型图 L^2 , 经全连接网络后就都变成了一个长度为 256 的特征矢量.最后用 Softmax 方法处理这些矢量,计算每种矢量属于逾渗还是非逾渗相的概率,选择概率大的作为对输入的预测即可.在整个网络中,所有神经元的传输函数都是限制线性单元(ReLU).并且为了防止出现过拟合的现象,在全连接网络中使用了 Dropout 正则化方法,让全连接网络中的神经元有 50% 的概率处于静息状态.训练过程中使用的是交叉熵作为代价函数,并使用 Adam 随机梯度下降法对网络进行整体训练.整个网络都是在基于 Python 语言下的 Tensorflow 库进行编写的.

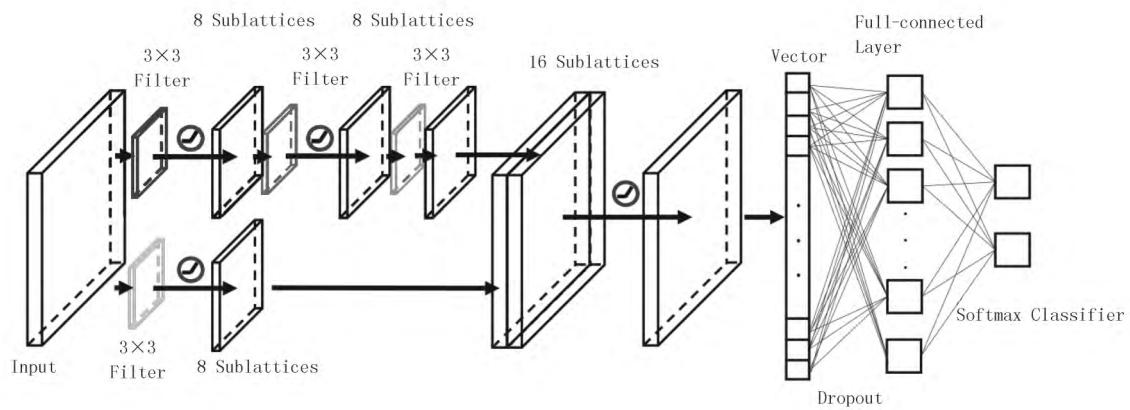


图1 卷积神经网络的结构

Fig.1 The architecture of the convolutional neural networks

除了使用上述的卷积神经网络外,还建立了一个新的全连接网络 GVM^[19].GVM 是全连接的前馈神经网络,如图 2.若设输入层隐含层和输出层的神经元数目为: N, M 和 L , 则网络的演化遵循以下规律.

对于隐含层,

$$y_i = f_i(\beta_i \bar{h}_i), \bar{h}_i = \sum_{j=1}^M \omega_{ij} x_j - b_i.$$

而对于输出层,

$$y_l = h_l, h_l = \sum_{i=1}^N \omega_{li} y_i.$$

其中, $y_i, f_i, \bar{h}_i, \beta_i$ 和 b_i 分别表示隐含层第 i 个神经元的输出,传输函数,局域场,传输函数常数和偏置. ω_{ij} 是输入和隐含层之间的权重,而 ω_{li} 则为隐含层和输出层之间的权重.

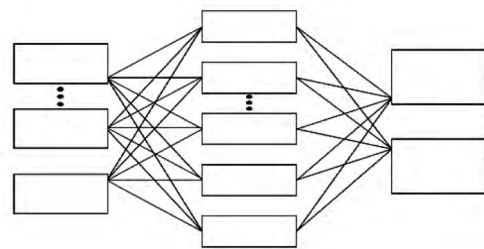


图2 一般向量机的结构

Fig.2 The architecture of the general vector machine.

对于神经网络参数的优化,GVM 采用 Monte-Carlo 算法.1) 随机地设 ω_{li} 的值为 ± 1 .2) 随机地选取其他的参数.3) 用这些参数计算出局域场 \bar{h}_i^q 和 h_l^q 以及相应的损失函数.4) 在参数中随机选取其中一类的某一个,在有效范围内做微小改变 ϵ . 5) 重新计算损失函数,如果新计算的损失函数与原先的损失函数相比更好或者相等,则保留这次改变,否则放弃这次改变.参数在演化过程中虽是任意选取但是仍有限制范围:

$$\beta_i \in [-c_\beta, c_\beta], \omega_{ij} \in [-c_w, c_w], b_i \in [-c_b, c_b].$$

需要根据训练集中不同的实际情况对 c_β, c_w 和 c_b 进行调整.这样才能达到学习机最好的学习效果.本文使用的 GVM 结构大致可以看作,输入层有 $L \times L$ 个神经元,即把每个二维构型都展成一个长度为 $L \times L$ 的向量作为输入.接着隐含层有 100 个神经元,输出层为 GVM 中定义的标准分类器^[19].

3 结果和讨论

用在打开概率 p 的取值范围 $[0, 1]$ 上随机均匀采样产生的 10 000 个构型作为训练集分别训练 2 个不同的学习机,再使用同样均匀采样的 10 000 个构型作为测试集,考察学习机的整体识别率.为重现逾渗概率曲线,又在区间 $[0, 1]$ 等间距地取 500 个参数 p ,在每个 p 下分别产生 10 000 个构型,构成测试集,用训练好的学习机对这 500 个测试集中的每个样本进行识别.得到了 3 组不同尺寸 $L = 32, 64$ 和 128 下机器预测的逾渗概率曲线和正确率曲线.作为对比,在计算预测的逾渗概率曲线时,对测试集样本通过模拟得出的标签进行统计,进而得到模拟曲线,如图 3(a)和图 4(a)的虚线所示.

图 3(a)和图 4(a)分别展示了 CNN 和 GVM 预测的逾渗概率曲线和模拟曲线.对于 CNN,预测的曲线和模拟得出的曲线在非相变区重合得很好而在相变点附近出现了偏差,而对于 GVM 而言,2 条曲线都重合得很好.说明在预测逾渗问题上,CNN 的表现与 GVM 相比略差一点.而且,2 种网络的预测曲线随尺寸增大,它们的转变区域也逐渐变窄.这是在有限尺寸效应的体现.另外,随着尺寸的增加,机器学习的表现也越来越好.对应的整体正确率,在 CNN 中分别为 96.2%,98.3%和 99.1%,在 GVM 中分别为 96.2%,98.3%,98.7%.在这种情况下,2 种学习机确实能以高正确率对逾渗构型进行识别,为判断逾渗提供了新的方法.

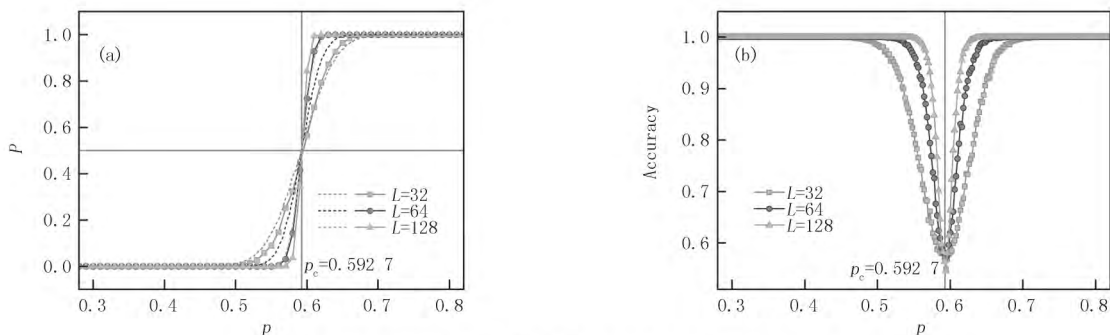


图 3 卷积神经网络的结果. (a), (b) 分别表示逾渗概率 P 和正确率与打开概率 p 的关系

Fig.3 The results of the convolutional neural network. (a), (b) are the relationships of percolation probability P and the accuracy with the open probability p of a site, respectively

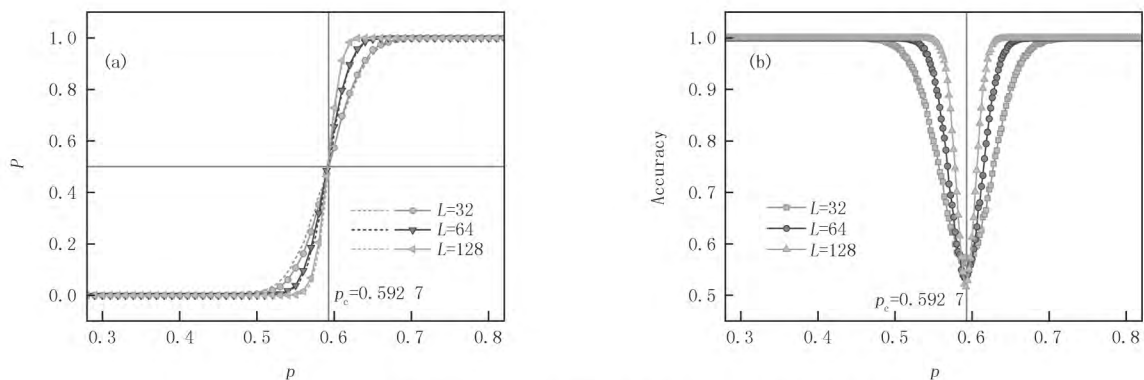


图 4 一般向量机的结果. (a), (b) 分别表示逾渗概率 P 和正确率与打开概率 p 的关系

Fig.4 The results of the general vector machine. (a), (b) are the relationships of percolation probability P and the accuracy with the open probability p of a site, respectively

图 3(b)和图 4(b)可进一步体现学习机对具体参数 p 下的构型的识别正确率的差异.有趣的是,正确率随着 p 接近真实相变点 p_c 越来越低,最低值在 p_c 处取得,然后越过相变点后,它又以镜像对称的方式又逐渐升高.不同尺寸下分别对应正确率为 56.2%,53.2% 和 51.6%,几乎是相当于随机猜测的正确率.这反映了相变问题中的一个普遍性质:临界慢化.越接近相变点,学习机的学习效果越差,并且学习效果最差的地方就是相变点 p_c 处.利用这个性质,可以定出相变点大约在 0.592 到 0.594 之间.这与传统方法算得相变点 0.592 7 相符.这是相变在机器学习角度的体现.另外,注意到在 Ising 模型中这样的正确率降低的显著性随着

尺寸增大而减小,逾渗模型正好与之相反^[3].这也从新的角度证明了逾渗模型在本质上同 Ising 模型的差异.

进一步地,对临界点附近的衰减行为做了定量的研究.传统的相变理论有一系列的关于临界行为的临界指数理论.在这里,如果把正确率 η 当作一个新的描述这类逾渗相变的物理量,就可以利用传统的临界指数理论来研究它.根据本文的结果,正确率关于相变点两侧对称,因此假定,正确率的临界行为也服从幂律:

$$\eta \sim |p - p_c|^\phi, \tag{1}$$

ϕ 就是正确率的临界指数.

为了计算 ϕ 的值,将正确率按照其关于相变点对称的性质分成 2 段: $p > p_c$ 的部分和 $p < p_c$ 的部分.因此对于每个尺寸上的数据,都有 2 段临界行为来定临界指数.并且,由于考虑到非常接近相变点时,正确率的计算误差会增加,因此为了避免这些误差带来的影响,不将这一部分数据计算在内.而幂律关系又只在临近相变点附近才会出现,所以离相变点非常远的那部分数据也不会被算在内.最终选取中间一段的数据点来定正确率的临界指数.具体方案是,直接利用之前的 CNN 和 GVM 的测试正确率的数据,将其坐标转化为双对数坐标,根据幂律关系式(1)可知,此时图像的斜率就是临界指数 ϕ 的值.只需要对数据点进行最小二乘法拟合求斜率即可.最终结果如图 5 所示.

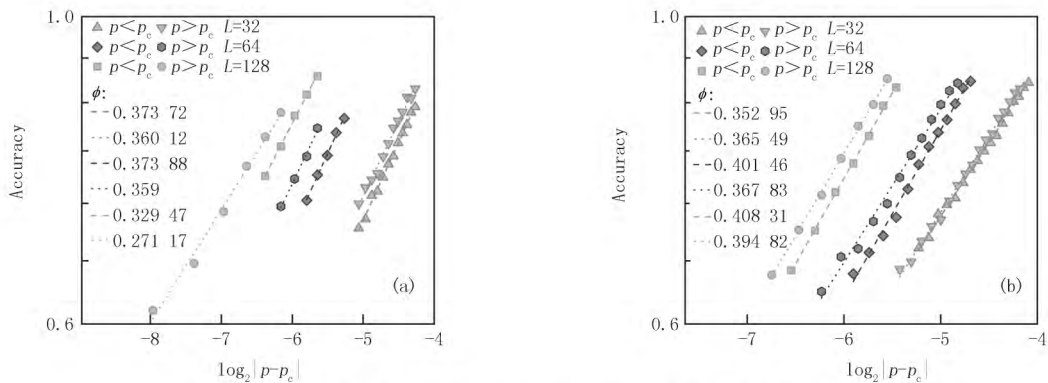


图 5 正确率与打开概率 p 的标度关系. (a), (b) 分别表示卷积神经网络和一般向量机的结果
Fig.5 The scaling relations between the accuracy and the open probability p . (a), (b) are the results of the convolutional neural network and general vector machine

算得在 CNN 的结果中, $\phi \approx 0.37$; 在 GVM 中 $\phi \approx 0.39$.在计算误差允许的范围内,在这 2 种网络中的 ϕ 可以认为是相等的.从图 5 中还可以看到,在不同尺寸下 ϕ 值也基本上相等,所以这个临界指数是一个与尺寸无关的量.这也同临界指数的定义相符合.那么也就是说,使用不同的机器学习方法计算的正确率虽然在数值上有所差异(这是不同的网络的表达能力的差异导致的),但是只要其正确完成了学习,那么它们正确率向临界点的趋近的定量行为是一致的.这与机器学习方法无关,是这个逾渗模型内禀性质,体现的是相变点奇异的普适性.而且,通过将 ϕ 值与已有的逾渗临界指数作对比,并没有发现与其一致的参数.因而,可以把机器学习上的正确率作为一个新的量,来描述这类相变.这也为研究这类逾渗问题提供了新的角度.

4 结 论

本文利用监督学习方法,通过训练卷积神经网络和一般向量机,对二维方格子上的座逾渗相变问题进行了研究.首先证明了机器学习对于逾渗这类不存在哈密顿量控制的动力学演化的相变模型,也可以通过监督学习的方式,正确被训练以及正确识别不同参数下对应的不同的相.区别于类 Ising 模型,机器学习在这类相变模型上的成功更具有一般性的意义.为进一步使用机器学习研究这类相变提供了实践基础.在此基础上发现,学习机的正确率会在相变点附近发生衰减,越过相变点后又会逐渐恢复.这个性质可以用于准确确定相变点的位置.计算了相变点附近不同尺寸下正确率随开放概率 p 的幂律衰减指数,发现在不同学习机中,该指数是普适的.这个结果从全新的角度展现了相变点奇异的普适性,同时也为相变研究提供了正确率这一全新参量.总之,本文研究为相变研究提供了新的角度,也为机器学习在物理上的应用提供了更加有力的证据.

参 考 文 献

- [1] Curtarolo S, Morgan D, Persson K, et al. Predicting crystal structures with data mining of quantum calculations[J]. *Physical Review Letters*, 2003, 91(13): 135503.
- [2] Carleo G, Troyer M. Solving the quantum many-body problem with artificial neural networks[J]. *Science*, 2016, 355(6325): 602-606.
- [3] Carrasquilla J, Stoudenmire M, Melko R. Machine learning phases of matter[J]. *Nature Physics*, 2017, 13: 431.
- [4] Portman N, Tamblyn I. Sampling algorithms for validation of supervised learning models for Ising-like systems[J]. *Journal of Computational Physics*, 2016, 350: 871-890.
- [5] van Nieuwenburg E P L, Liu Y H, Huber S D. Learning phase transitions by confusion[J]. *Nature Physics*, 2017, 13(5): 435.
- [6] Zhang W, Liu J, Wei T C. Machine learning of phase transitions in the percolation and XY models[EB/OL]. [2018-11-04]. <https://arxiv.org/abs/1804.02709>.
- [7] Wang L. Discovering phase transitions with unsupervised learning[J]. *Physical Review B*, 2016, 94(19): 195105.
- [8] Zhang Y, Kim E A. Quantum loop topography for machine learning[J]. *Physical Review Letters*, 2017, 118(21): 216401.
- [9] Broadbent S R, Hammersley J M. Percolation processes: I. Crystals and mazes[J]. *Mathematical Proceedings of the Cambridge Philosophical Society*, 1957, 53(3): 629-641.
- [10] Isichenko M B. Percolation, statistical topography, and transport in random media[J]. *Reviews of Modern Physics*, 1992, 64(4): 961.
- [11] Basta M, Picciarelli V, Stella R. An introduction to percolation[J]. *European Journal of Physics*, 1994, 15(3): 97.
- [12] Beer T, Enting I G. Fire spread and percolation modelling[J]. *Mathematical and Computer Modelling*, 1990, 13(11): 77-96.
- [13] Grimmett G. What is percolation? [M] // *Percolation*. Berlin: Springer, 1999: 1-31.
- [14] Batrouni G G, Hansen A, Larson B. Current distribution in the three-dimensional random resistor network at the percolation threshold [J]. *Physical Review E*, 1996, 53(3): 2292.
- [15] Newman M E J, Ziff R M. Fast Monte Carlo algorithm for site or bond percolation[J]. *Physical Review E*, 2001, 64(1): 016706.
- [16] Stockman G, Shapiro L G. *Computer vision* [M]. Upper Saddle River: Prentice Hall PTR, 2001.
- [17] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436.
- [18] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C] // *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE Press, 2016: 2818-2826.
- [19] Zhao H. General vector machine [EB/OL]. [2018-11-04]. <https://arxiv.org/abs/1602.03950>.

Study site percolation phase transitions based on machine learning

Xu Rongxing, Zhao Hong

(Department of Physics, Xiamen University, Xiamen 361005, China)

Abstract: Different from traditional physical approaches, machine learning technics are widely applied in various fields benefitted from their high accuracy and efficiency in detection and classification of complex data sets. Identifying phase transitions are the most typical works in the field of combination of machine learning and statistical physics. So far, researches are mainly focused on the spin models with dynamical evolutions, such as Ising model. However, a more general type of phase transition models, such as the percolation model, which is not determined by dynamical evolutions but the intrinsic property of itself, has not been researched yet. In this article, combined with the new machine learning methods, the convolutional neural network and the general vector machine, we study the two-dimensional square-lattice site percolation model. We find that well-designed networks can identify different configurations with a high accuracy, which proves that machine learning works on this problem. Then, by calculating accuracies of our networks for identifying configurations on different parameters, we find the accuracy drops near the critical point in both models, and this decay follows the power-law with the parameter, which is corresponding to the traditional phase transition theory. We calculated the exponents of the power-law and find the exponents are nearly the same in both different models. Our research does not only reveal the universality of phase transitions, but also provides a new way to find the critical point.

Keywords: machine learning; convolutional neural network; general vector machine; site percolation; phase transition; accuracy; scaling relation

[责任编辑 杨浦]