



中南大學

CENTRAL SOUTH UNIVERSITY

本科毕业设计 (论文)

GRADUATION DESIGN (THESIS)

题目: 机器学习方法在强关联系统中的应用

学生姓名: 王梦凡

指导教师: 杨义峰 研究员 中国科学院物理研究所

易有根 教授 中南大学

学院: 中南大学物理与电子学院

专业班级: 应用物理 (T)

本科生院制

2018年6月

摘 要

在强关联体系中，电子之间具有较强的关联相互作用，无法采用传统的微扰理论对其微观体系进行计算，因此，一系列模型和数值计算方法被构建出来进一步研究强关联体系。对于有限的格点来说，以上思路能够有效地算出其哈密顿量，但随着格点的增多，**指数墙**困难不可避免会发生，针对当前算力的情况下，要计算大量格点还存在着许多困难。与此同时，机器学习方法以其能够跨越指数墙难题的特点引起关注，结合机器学习方法的强关联体系计算能够有效地解决物理上一些人力难以解决的问题。

本文主要分为三个部分。第一部分主要是介绍强关联电子体系，分为模型，方法，应用三部分。模型主要介绍了强关联系统在不同情况下使用的三种模型：**Hubbard 模型**，**近藤模型**和 **Anderson 模型**，方法部分介绍了**严格对角化**，**密度矩阵重整化群**和**数值重整化群**等方法以及利用这些方法求解相应模型的过程。第二部分主要介绍了机器学习的相关构架与模型方法。第三部分将机器学习和强关联体系相结合，主要研究了机器学习区分物理**相变**的过程，引入二维方格子 **Ising 模型**，通过机器学习得到相变温度临界点，与理论计算所得的结果一致。此外，还简要介绍了其在利用神经网络求解量子波函数和基态能量的过程。

无论是机器学习还是强关联体系都包含着丰富的内容，其都经过了长久而又较为系统的发展过程。目前其方法的结合交叉还较为表面，更加深层次的推理，更进一步的结合还具有很大的发展空间，相信在未来更深层次的交叉发展能够使强关联体系和机器学习领域都得到较大发展。

关键词： 强关联体系，机器学习，指数墙，相变，波函数

Abstract

Strong correlation between electrons exists in strongly correlated electronic systems, and it is not possible to use conventional perturbation theory to calculate the microscopic system. For this phenomenon, a series of models and numerical methods have been developed and used. For limited lattice points, the above algorithm can effectively calculate the Hamiltonian, but as lattice points increasing, it will inevitably result in the difficulty of **the Exponential Wall**, for the limits of the current calculation force. There are still many difficulties to solve in order to calculate a large number of lattices. At the same time, the machine learning method attracts attention because of its ability to cross the exponential wall. Combining the calculation of the strong association system under the machine learning method can effectively solve some complex physical problems.

This article is divided into three parts. The first part mainly introduces the strong associative electronic system, which can also be divided into three parts: models, methods and applications. The model mainly introduces three models designed for strong association system in different situations: **Hubbard model, Kondo model, and Anderson model**. The method section mainly introduces methods such as **Exact Diagonalization, Numerical Renormalization and Density Matrix Renormalization**, and solving the relevant model with these models. The second part mainly introduces the related framework and model method of machine learning. The third part combines machine learning and strong correlation systems. It mainly studies that machine learning distinguishes the phase transition process in physics. With the two-dimensional Ising model, the **phase transition** temperature point is obtained by machine learning, which is consistent with the theoretically calculated results.

Whether it is machine learning or strong correlated system, it contains rich content, and undergoes a long and systematic development process. At present, the combination of its methods is still relatively superficial, deeper studying, and the combination also has a lot of room for development. It is believed that further combined development will enable the development of strong associative systems and machine learning.

Keywords: Strong Correlated System, Machine Learning, Phase Transition

目 录

摘 要	I
Abstract	III
第 1 章 绪论	1
第 2 章 强关联体系	3
2.1 强关联电子体系的基本模型	3
2.1.1 Hubbard 模型	3
2.1.2 近藤模型	4
2.1.3 周期性 Anderson 模型	4
2.2 强关联电子的计算方法	5
2.2.1 ED	5
2.2.2 NRG	5
2.2.3 DMRG	8
2.3 强关联电子的应用	9
第 3 章 机器学习	11
3.1 机器学习体系	11
3.2 机器学习的模型	14
3.2.1 神经元模型	14
3.2.2 受限玻尔兹曼机	15
3.2.3 卷积神经网络	17
3.3 机器学习方法	17
3.3.1 PCA	18
3.3.2 逆向误差传播算法	18
3.3.3 核方法	19
第 4 章 机器学习在强关联体系中的应用	21
4.1 对相的分类	21
4.2 研究波函数与基态能量	25
第 5 章 当前进展	29
第 6 章 未来展望	31
参考文献	33

致谢	35
附录 A: 程序代码	37
附录 B: 英文文献翻译	39
附录 C: 英文文献原文	41

第1章 绪论

物理研究的问题是多粒子组成的各种系统，其中凝聚态方向是其中的主流。凝聚态物理主要是从微观结构出发，对无限复杂的电子、原子核、原子、磁矩、量子比特等组合的集体行为研究，其研究的方向包括凝聚态的结构，动力学原理及其宏观物理性质之间的联系。以此为依据，其包含的固体物理分支中的强关联体系由于高温超导、量子霍尔效应、超冷原子物理、量子磁性等前沿领域近年来的发展受到广泛关注。

强关联电子体系区别于之前利用单电子能带理论近似求解的无关联或者弱关联体系，其主要体现在过渡金属及其化合物，镧系（或铜系）元素及其化合物之间，其共同特征在于其具有部分填满的 $3d$ 、 $4d$ 、 $4f$ 或者是 $5f$ 电子壳层，电子结构的独特结构使之具有十分新奇的物理现象，比如说金属-绝缘体相变。上个世纪三十年代，量子力学的引入解释了凝聚态理论中铁磁模型和自旋模型的奇异相变，于此同时，超导和近藤效应却亟待解决，之后，量子场论、朗道准粒子、临界现象、平均场理论、量子霍尔效应、高温超导、BCS 理论、拓扑绝缘体等诸多理论随后发展，固体物理，凝聚态物理，量子物理理论体系逐步完善，其为强关联系统的发展奠定了一定的根基。

理论物理包括使用物理模型来计算理解物理状态的性质，其中包括研究固体电子性质的模型，如 Band 结构，Drude 模型和密度泛函理论。还有研究物理相变的模型，例如临界指数，Ginzburg-Landau 理论还有数学方法量子场论以及重整化群的使用。现代理论研究一般使用相关的数值计算和部分数学工具来理解诸如高温超导性，拓扑相位和规范对称性等现象。理论计算晶体的性质多依赖于其布洛赫波函数，在计算多体波函数中，由于金属电子之间具有复杂的相互作用，精确计算很难，因此需要近似的方法来进行有意义的预测。近似的方法主要包含：变分法、密度泛函理论、平均场理论等。

一般的量子力学多体问题被普遍认为是 NP 完全的，即在多项式时间里无法找到合适的算法来得出多体问题的结果，通常以“指数墙”来描述这种计算时间随着计算复杂度指数增加的问题。在现有的计算机水平中，在计算宏观物体，比如说一滴水的时候（其中约有一万四千亿亿个水分子），根据经典物理以及量子物理相关的法则，依据薛定谔方程，原则上可以求解出其能量、波函数、粘滞系数等相关的性质。但在构造计算波矢的时候，以计算自旋为 $\{\uparrow, \downarrow\}$ 为例，假设有 100

个粒子，其波矢空间大小为 10^{30} 量级大小，按照现有的计算机速度，要计算一滴水的时间大约为： 10^{25} 年，而且当前计算机的内存也无法满足这么大量的计算。

针对这一问题，物理学家在过去的几十年里，发展了一系列方法来解决这一问题，如**密度重整化群、动力学平均场理论、量子蒙特卡洛、严格对角化和张量网络方法**等等，但针对不同的问题各有限制，尚无法得出一个令所有人为之满意的结果。

与此同时，机器学习的方法在这十年中步入人们的视野。**机器学习是计算机训练大量的数据来找到其中的规律以及宏观属性的一种方法。它能够越过指数困难，直接找到数据的内在属性^[1]**。同物理问题是从大量微观粒子中预测和解释宏观现象的本质特征对比，机器学习训练计算机从大量数据中探测和找到数据的宏观特征与此同源。

机器学习主要可以分为三个大类：**监督学习，无监督学习和强化学习**，通过带有标签或者不带标签的数据集找到数据集中的规律，从而对数据进行归一或者分类。利用这种性质，可以找到物理中不同相的临界点，优化量子多体问题中的算法，可以拟合求解波函数和其能量等等。与此同时，物理学科的也可以对机器学习领域的发展提供一定的思路，将机器学习方法引入到物理中来，能够促进学科之间的相互交叉，相互发展。

第2章 强关联体系

强关联体系 (Strongly Correlated Systems, SCS) 是指系统中粒子拥有部分填满的 d 或 f 电子壳层, 同时保有局部磁矩, 这种具有强电子—电子相互作用的系统都属于强关联电子体系, 简称强关联体系。电子的性质决定物质的结构, 传统的单电子能带论在金属和半导体上取得了一定的成就, 因为这类材料中的电子之间的相互作用比较弱, 可以忽略。但在强关联电子体系中, 电子之间的相互作用已经达到电子能带宽度的量级, 不可忽略。因此, 需要新的理论方法来计算和解释这类材料的奇异现象。

2.1 强关联电子体系的基本模型

强关联电子的基本计算方法主要有两类, 一类是以真实材料的参数为输入, 计算材料的电子结构信息, 被称为第一性原理计算。另一类是给出一类模型的哈密顿量, 据此来求解结果, 一般情况下只有数值解。对包含未填满的 d 或者 f 的强关联电子体系, 根据电子的局域程度和跃迁关系, 强关联理论中主要包含以下几种模型。

2.1.1 Hubbard 模型

Hubbard 模型是研究金属中具有强相互作用行为电子的重要模型, 此模型哈密顿量为:

$$H_H = -t \sum_{ij\sigma} C_{i,\sigma}^\dagger C_{j\sigma} + U \sum_j n_{j\uparrow} n_{j\downarrow} \quad (2.1)$$

其中 i, j 是格点的索引, σ 是自旋索引, t 是临近电子之间跳跃的重叠能量, n 为电子占据数, C^\dagger, C 为电子的产生湮灭算符, U 为库仑相互作用参数。Hubbard 哈密顿量的第一项参数是跃迁项, 即动能项, 是用来描述相邻格点之间的跃迁, 第二项为局域库仑排斥项, 描述的是在同一个格点上具有相反自旋的电子之间的相互作用。

Hubbard 模型不同于以往单电子能带模型的主要一点是: 单电子理论认为价电子在晶体中的运动是彼此独立的, 电子之间的关联作用被忽略了。但 Hubbard 模型考虑到当电子由一个轨道跃迁到另一个时, 须得考虑后一轨道是否已被其他电子所占据, 如果已被占据, 则应该计入同一原子 (或格点) 周围两个电子之间的库仑作用, 由此体现强关联作用。近年来, Hubbard 模型广泛应用于强关联

电子体系，发展了著名的 $t - J$ 模型，用于建立铜氧化合物的高温超导理论。此外，Hubbard 模型还用于讨论金属磁性问题，并在 20 世纪 80 年代初期发展了巡游电子磁性自旋涨落理论。

2.1.2 近藤模型

在非磁性金属中掺入磁性杂质形成稀磁合金，并在低温下出现电阻极小的现象。在非磁杂质的剩余电阻与温度无关的情况下，随温度降低磁性杂质对电阻贡献率增大。1964 年近藤 (Kondo) 通过 $s - d$ 模型解释了当电子被磁杂质散射的时候，不仅电子的自旋变化，磁杂质的自旋也在变化。通过对此散射问题的微扰计算，求出杂质电阻随温度降低呈现指数增长，结合电子被热声子散射的电阻，解释了电阻在非零点极小的现象。除电阻的反常之外，低温下，比热容，磁化率等一系列的反常现象，统称为近藤效应。描述近藤效应的哈密顿量是：

$$H = \sum_{k\sigma} \epsilon_k C_{k\sigma}^\dagger C_{k\sigma} - \frac{J}{N} \sum_{jkp} e^{iR_j \cdot (k-p)} [(C_{k\uparrow}^\dagger C_{p\uparrow} - C_{k\downarrow}^\dagger C_{p\downarrow}) S_j^{(z)}] + C_{k\uparrow}^\dagger C_{p\downarrow} S_j^{(-)} + C_{k\downarrow}^\dagger C_{p\uparrow} S_j^{(+)} \quad (2.2)$$

第一项代表自由电子的动能项，其他项代表在格点 R_j 的局域自旋的散射态。

2.1.3 周期性 Anderson 模型

周期性 Anderson 模型有两种电子：巡游电子 (s 电子) 和局域电子 (f 电子)，Anderson 模型中巡游电子和局域电子通过杂化项耦合，其哈密顿量是：

$$H = \sum_{ij,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \epsilon_d \sum_{i\sigma} n_{i\sigma}^d + U \sum_i n_{i\uparrow}^d n_{i\downarrow}^d + \sum_{ij,\sigma} (V_{ij} c_{i\sigma}^\dagger d_{j\sigma} + h.c). \quad (2.3)$$

其中 $c^\dagger(c)$ 和 $d^\dagger(d)$ 分别为 s 电子和 d 电子的产生 (湮灭) 算符， $n_{i\sigma}^d$ 是自旋为 σ 的 d 电子在 i 格点上的占据数， ϵ_d 是电子的能级， V_{ij} 是杂化参数，描述 s 和 d 电子之间的相互关系。Anderson 模型主要用于解释 Kondo 绝缘体、局域磁矩的形成、重费米子化合物等等。Anderson 模型和近藤模型都描述了连续电子与局域电子的相互作用。将 Anderson 模型进行一个规范变换，可以变成一个类似于近藤的模型。Anderson 模型和 Hubbard 模型是强关联电子体系的主要基础模型，理解上述的几个模型可以于加深我们对金属-绝缘体相变、重费米子效应、高温超导等强关联电子体系等物理认知。

以上三个模型有一个共性，即都只能得出严格的一维解，但对于更高维的来说，模型无法求解，因此并不能普遍推广^[2]。

2.2 强关联电子的计算方法

在物理计算领域，由于研究的物理世界存在着复杂的相互作用以及各种奇异现象，现有的数学能力很难在高维上求解其解析解，因此，在物理上发展了多种关于物理实际问题的数值方法，以下主要概括几个物理学中比较常用的强关联计算方法。

2.2.1 ED

严格对角化（Exact Diagonalization Method）方法应用在计算有限格点的哈密顿量上面。在计算体系的哈密顿量的时候，其流程是：选定相应的表象，构造基矢，将哈密顿量在相应的基矢下展开，求解得到相应的哈密顿矩阵，之后对矩阵进行对角化操作，求出相应的本征值跟本征矩阵。

以一维 **Heisenberg** 模型为例，其哈密顿量为：

$$H = -J \sum_i \hat{S}_i \cdot \hat{S}_{i+1} \quad (2.4)$$

引入升降算符：

$$\hat{S}_j^{\dagger} = \hat{S}_x + i\hat{S}_y \quad \hat{S}_j^{-} = \hat{S}_x - i\hat{S}_y$$

且自旋的希尔伯特空间为： $|\psi_k\rangle_{L=N} = \underbrace{\{|\uparrow\rangle, |\downarrow\rangle\} \otimes \{|\uparrow\rangle, |\downarrow\rangle\} \otimes \{|\uparrow\rangle, |\downarrow\rangle\} \otimes \cdots}_{N}$ ，将哈密顿量在希尔伯特空间上展开作用，可以得到哈密顿的矩阵，具体公式为：

$$H_{ij} = \langle \psi_i | \hat{H} | \psi_j \rangle$$

可以据此求解哈密顿的本征值和本征矢。

这种方法在研究少量格点的哈密顿的时候方便且简洁，但由于在构造基矢时，对于 N 个格点，其希尔伯特空间的大小为： $2^N \times 2^N$ ，大小呈指数形式上升，存在着指数墙的困难。因此，在直接使用严格对角化的方法的时候，存在着较大的局限性，仅能计算少量的格点。

2.2.2 NRG

数值重整化群（Numerical Renormalization Group）这种方法主要在有杂质引入的格点模型中使用，其主要操作是指对体系进行不断地重整化操作，将一些无关的参数消掉，保留有用的参数^[3]。

以单杂质 **Anderson** 模型为例，其模型的哈密顿量为：

$$H = H_{imp} + H_{bath} + H_{imp-bath}. \quad (2.5)$$

其中,

$$H_{imp} = \sum_{\sigma} \varepsilon_f f_{\sigma}^{\dagger} f_{\sigma} + U f_{\uparrow}^{\dagger} f_{\uparrow} f_{\downarrow}^{\dagger} f_{\downarrow}$$

$$H_{bath} = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^{\dagger} c_{k\sigma}$$

$$H_{imp-bath} = \sum_{k\sigma} V_k (f_{\sigma}^{\dagger} c_{k\sigma} + c_{k\sigma}^{\dagger} f_{\sigma})$$

在利用 NRG 方法计算多个格点的哈密顿量时, 其操作步骤为:

1. 将哈密顿量变换到一维能量表象下;
2. 对数离散化能量区间;
3. 将离散的能量映射到一个半无限长的链上 (链的第一个格点为杂质部分, 剩余的半无限链为格点动能);
4. 迭代对角化求解链的本征值和本征能量。

在以上的迭代过程中, 同样存在着 ED 方法所面临的指数墙困难, 针对这一问题, Wilson 通过逐步构造求解链的方法, 即: 在单杂质的基础上, 逐步增加格点, 同时利用截断除掉体系的高能态, 在低能态的基础上生成新的哈密顿量, 能在保证计算精度的前提下, 计算较多格点。其流程图如图 2.1。

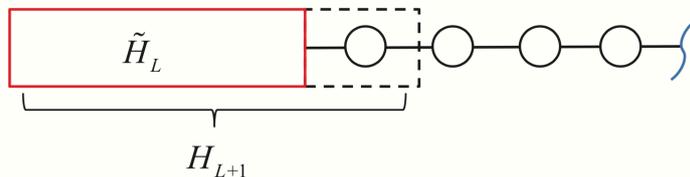


图 2.1 NRG 算法过程图示^[4]

根据 NRG 算法的具体流程, 计算单杂质 Anderson 模型的谱函数和自能函数, 结果如图 2.2。(相关的程序见附录 A)

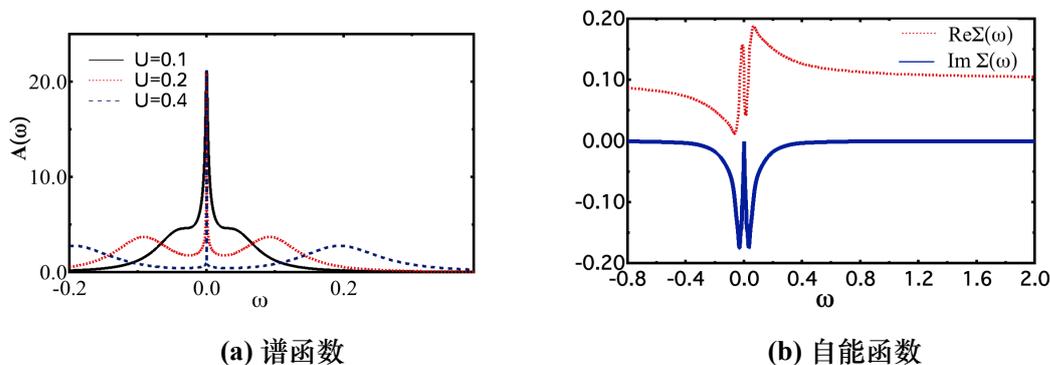


图 2.2 NRG 计算单杂质 Anderson 模型的谱函数和自能函数。

根据图 2.2，图 (a) 谱函数说明随着初始能带宽度的增大，其体系由金属态逐渐转变为绝缘态。图 (b) 中其实部为奇对称，虚部为偶对称，且实部在原点关于 ω 的导数为负，且很大，说明其有效质量很大。

NRG 方法在处理单杂质模型的时候能够取得良好的效果，但由于其基本方法是截断体系的高能态，在处理量子晶格问题上，由于其边界低能部分为零，绝大部分为高能态，因此，此方法不再适用。

在一般的物理计算中，考虑到有杂质引入的模型并针对相互作用的量子系统，经常结合动力学平均场理论和数值重整化群一起计算。

平均场方法的本质是将晶格模型映射到一个有效平均背景场中的单点系统。示意图如图 2.3 所示。针对这一问题背景，采用 Hubbard 模型来计算其谱函数和

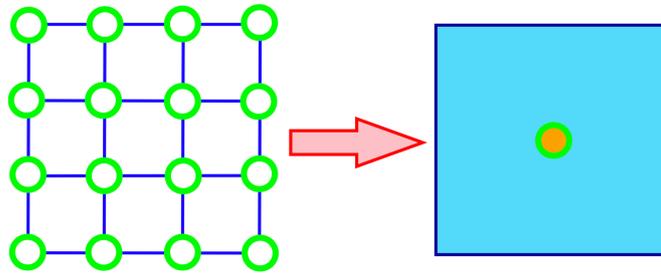


图 2.3 晶格模型映射为单点参考系统的示意图^[5]

自能函数，相应的结果如图 2.4 所示。

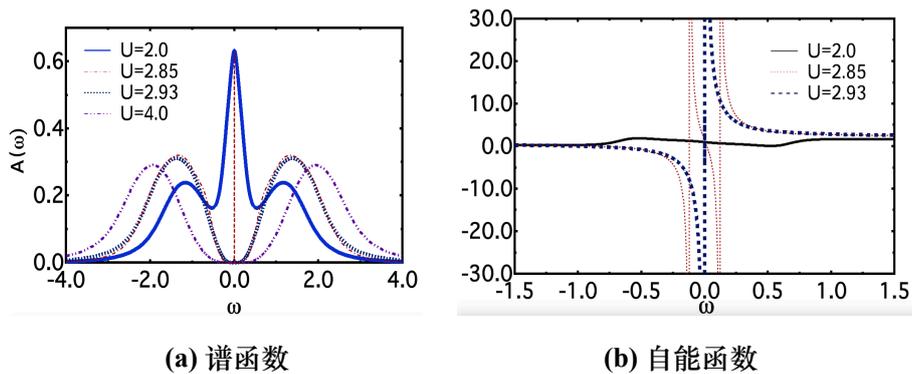


图 2.4 NRG-DMFT 计算 Hubbard 模型的谱函数和自能函数

NRG 算法的实现是重复 ref[3] 的工作，所得的结果不仅说明了其模型随着初始能带宽度的改变，而且得出其金属-绝缘体相变发生在 $U=2.93$ 时。

以上的方法能解析地解出单杂质 Anderson 模型和 Hubbard 模型的哈密顿量。是解决相关物理问题中不可或缺的重要方法。

2.2.3 DMRG

密度矩阵重整化群（Density Matrix Renormalization Group, DMRG）为了解决一维数值重整化群方法下其对边界问题失效的问题的重新改进。将体系分为系统块和环境块，如图 2.5 所示，初始格点只有系统与环境相连接的部分，之后通过不断地增加系统块和环境块，每增加一次格点，求解当前的约化密度矩阵和其本征值，本征函数，并通过适当地截断保证格点的不断增加，当达到指定的格点数时，求得整个系统的基态能量和波函数。

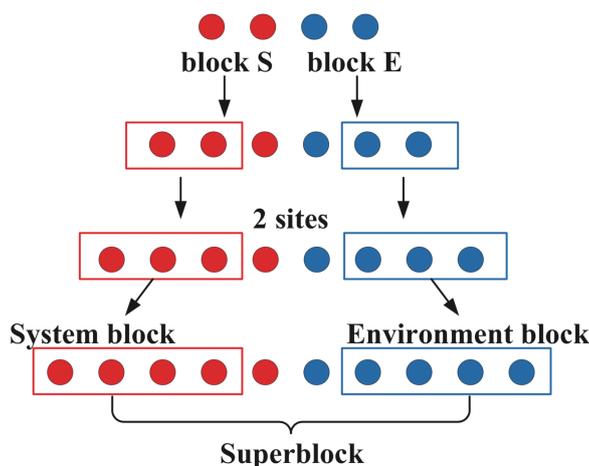


图 2.5 密度矩阵重整化群示意图

以近藤海森堡模型为例，其哈密顿量为：

$$H = -t \sum_{n=1}^{N-1} \sum_{\sigma=\pm} (c_{n,\sigma}^\dagger c_{n+1,\sigma} + h.c.) + J_k \sum_{n=1}^N s_n \cdot S_n + J_H \sum_{n=1}^{N-1} S_n \cdot S_{n+1} \quad (2.6)$$

式中第一项是导带电子的跃迁项， t 代表其跃迁参数。第二项是导带电子与局域自旋的耦合项，其中 J_k 是其耦合系数。第三项是局域自旋之间的相互耦合作用， J_H 是其耦合系数。

据此，引入导带电子电荷密度 n_i^c 、局域杂化 v_i 和局域自旋最近邻关联函数 χ_i 物理量，为：

$$n_i^c = \sum_{\sigma} \langle c_{i,\sigma}^\dagger c_{i,\sigma} \rangle, \quad v_i = \langle S_i \cdot s_i \rangle, \quad \chi_i = \langle S_i \cdot S_{i+1} \rangle \quad (2.7)$$

其中局域杂化代表的物理量反映了局域自旋与导带电子的杂化行为，在实空间的行为称为杂化波。针对研究近藤空穴的单独作用的目的，并消除边界的影

响，考虑系统掺杂近藤空穴前后的物理量的变化由此定义三个变量为：

$$\begin{aligned}\delta n_i^c &= \sum_{\sigma} (\langle c_{i,\sigma}^{\dagger} c_{i,\sigma} \rangle - \langle c_{i,\sigma}^{\dagger} c_{i,\sigma} \rangle_0) \\ \delta v_i &= \langle \vec{S}_i \bullet \vec{s}_i \rangle - \langle \vec{S}_i \bullet \vec{s}_i \rangle_0 \\ \delta \chi_i &= \langle \vec{S}_i \bullet \vec{S}_{i+1} \rangle - \langle \vec{S}_i \bullet \vec{S}_{i+1} \rangle_0\end{aligned}\quad (2.8)$$

根据这三个变量，其相应的计算图形如图 2.6，其图是师兄，胡丹青师兄和杨义峰老师文章^[6]部分工作的重复。

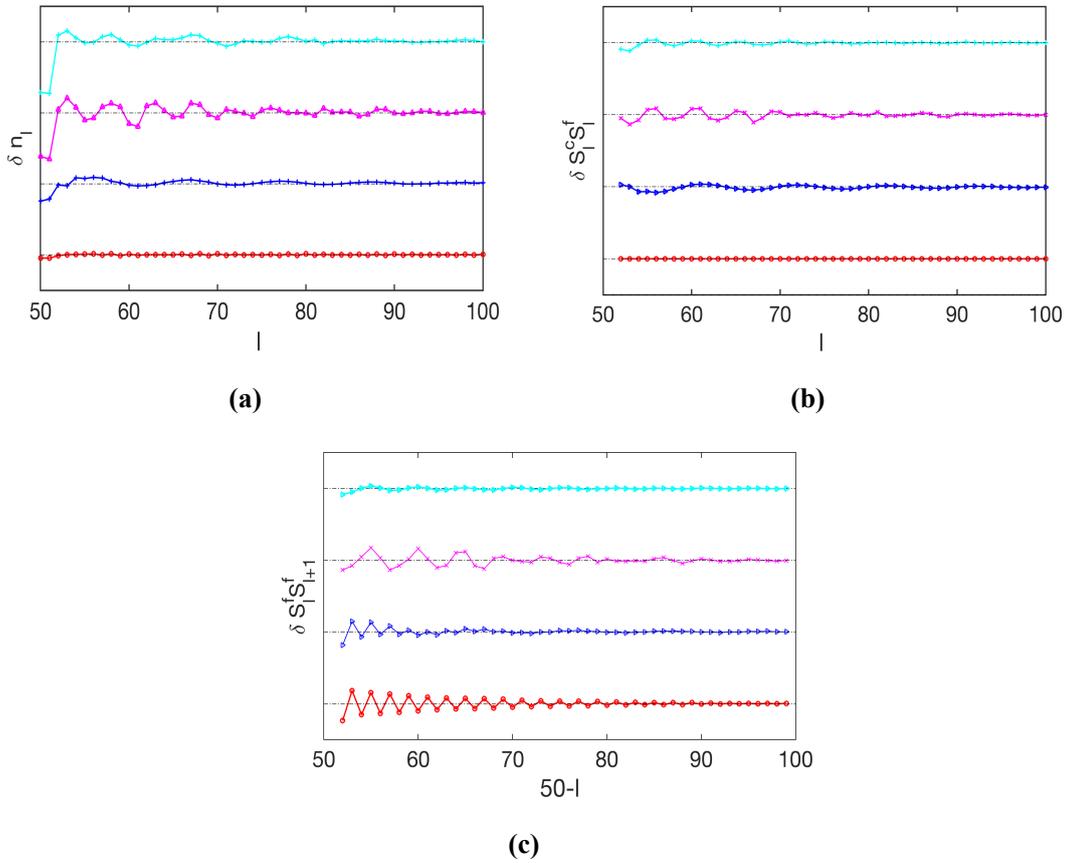


图 2.6 DMRG 在粒子占据数为 0.8，链长为 100 时的结果。

根据图可知，随着其耦合系数的增强，其杂化波震荡行为增强，且逐步显现一定的周期性。杂化波在格点中间呈现较强的震荡行为，并向两边界衰减。

2.3 强关联电子的应用

在凝聚态中电子之间相互作用不可忽略的强关联体系具有不可忽视的意义。在高温超导体、锰氧化物材料中的巨磁阻效应、分数量子霍尔效应、二维高迁移率材料中的金属-绝缘体相变重费米子系统、量子临界现象和量子相变、一维

导体中的电荷密度波等前沿研究问题都需要强关联的理论来解释。

此外，强关联体系中要研究各种粒子之间的强相互作用，采用的是还原论的方法，即：在利用微观粒子之间的基础相互作用，推广到其宏观产生的现象。微观粒子之间的作用推广到集体宏观的结果，不仅仅对解释物理世界的现象有重大作用，对于人类社会关系也有一定的参考意义。

第3章 机器学习

人工智能和神经网络经过几十年的发展，趋于完善。2017年 AlphaGo 以 3 比 0 的战绩打败当时围棋世界第一的人类棋手柯洁，一战成名，并引发世界研究人工智能的热潮。

实际上，人工智能在二十世纪五十年代（数字计算机尚在初级探索的年代）就已经被提出来了，最为出名的是计算机之父图灵于 1950 年介绍的「模仿游戏 (Imitation Game)」后称为「图灵测试 (Turing Test)」，即：让一个人与计算机对面的另一个个体进行交谈，并判断其是人还是计算机。本质上是尝试构建模仿成年人的机器的人工智能，让机器像人一样思考。人工智能发展至今，不到 70 年，但其进化速度远远超人类五千多万年进化的速度，不同的是人类依靠自然进化，而人工智能依靠人类进化，是人类智慧的结晶。人工智能 (AI) 的领域包括但不限于规划、问题求解、知识表征、搜索、自然语言处理、博弈、神经网络、专家系统、机器推断、统计机器学习、机器人、移动机器人、同时定位与地图创建、图像理解、计算机视觉等等。

3.1 机器学习体系

人工智能界有这样一则说法：机器学习是最能体现人工智能的一个分支，从其发展来看，其也是人工智能界发展最快的一个分支。现阶段，机器学习主要以统计学习为主要研究发展点，逐步探索扩充整个机器学习领域。根据机器学习的使用方法，其可以分为三方面：**监督学习**、**无监督学习**和**强化学习**。下面从这三个方面来介绍机器学习。

1. 监督学习 (Supervised Learning, SL)

监督学习一般是指利用一组已知类别的样本来调整分类器的参数，使之能够达到所要求的性能的学习过程。在这个过程中，将输入可能取值的空间定义为输入空间，输出可能取值的空间定义为输出空间，训练集中精确的输出层数据称为标签 (Labels)。监督学习的目的在于**学习一个可以表示输入到输出的映射的模型**^[7]。

根据模型中输入和输出变量的类型，其预测的任务中有不同名称。当输入和输出为连续变量时，其称为**回归问题**；当输出变量为有限离散变量时，称为**分类**

问题；当输入与输出为序列时，为**标注问题**，并可以据此将模型分为：条件概率分布 $P(Y|X)$ 和决策函数 $Y = f(X)$ 。

监督学习的流程操作是：

- (a) 确定训练集（包含特征值和标签值） $(\{x_n \in R^d, y_n \in R\}^N)$ ；
- (b) 利用训练集中输入层的数据计算输出层；
- (c) 与标签值对比计算误差，通过迭代找到最优模型参数。

所谓的机器学习即模型不断地从数据中学得其规律，自我优化，从而达到最优目的过程。

监督学习的算法有很多，常用的经典算法有：**邻近算法、逻辑回归、支持向量机、神经网络、朴素贝叶斯分类器、决策树、随机森林**等另外还有一些比较新的算法，比如**深度信念网络和胶囊网络**。当原始数据只有输入值，却没有已知的输出值标签时，问题就变为无监督学习了。

2. 无监督学习 (Unsupervised Learning, UL)

无监督学习与监督学习不同的地方在于：无监督学习的训练集没有标签，其通过训练未知类别的样本学得其数据结构。程序在此类学习方法中通过给定机器系统具有一定规律的数据集，然后通过数据集中数据之间的相关关系与对应的规律来找到某种共性特征，结构。

无监督学习可分成两大类：输入数据聚类和输入特征变量关联。其操作流程是：

- (a) 给定训练集（带有特征值，无标签） $(\{x_n \in R^d\}^N)$ ；
- (b) 根据相对应的算法求解数据集之间的相关关系；

在无监督学习中要求输入变量的独立性，机器学习中需要专门的工程通过对输入变量相关性检查从而来对数据进行预处理操作，进一步增加模型的计算精度和降低计算量。

常用的无监督学习算法有：**主成分分析算法、K 均值聚类、自组织映射神经网络、受限玻尔兹曼机**，另外还有一些利用物理中量子理论发展的算法，比如说**量子系综算法**等等。

监督学习与无监督学习的差别体现在**数据集中主要是有无标签的区别**，但在具体的实现与操作流程中，过程完全不同，能达到的效果，应用的实际情况也截然不同。监督学习根据其实现是根据学的模型判断未知的样本，可以用在医学中比如说机器自主判断 X 光照片中病人是否患有癌症，用在互联网中比如说对相关的新闻增加标签分类，用在日常生活比如对数字图片进行识别；无监督学习

根据数据集判断其中规律，比如说通过检测网络流量的异常模式发现其中的操作，图像识别与重建等等。针对数据其处理过程的不同见图 3.1。

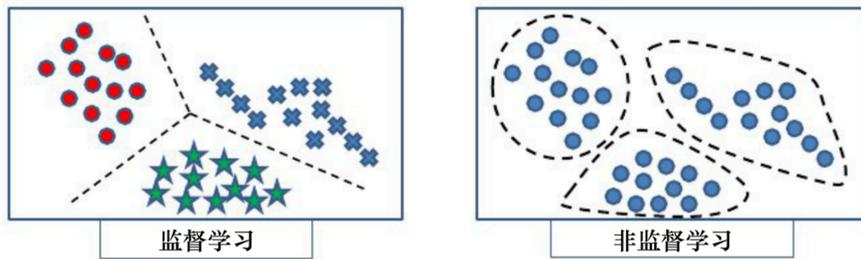


图 3.1 监督学习与无监督学习对数据集的不同作用

由图 3.1 可知，监督学习主要体现在对数据的分类，而无监督学习体现在对数据进行聚类，寻找共通集上。

3. 强化学习 (Reinforcement Learning, RL)

强化学习与监督学习和无监督学习不同，其过程中没有已知数据集来学习，且面对的是一个不断变化的状态空间，要解决的是一个决策链问题。其目的是找到当前状态下的最优决策，机器根据多次决策产生的大量试错学习来决定产生最佳结果的决策。其过程包含四个元素：环境、对象、交互、反馈，即通过对象与环境的交互行为产生的反馈来决定决策的决策问题。其示意图如 3.2。

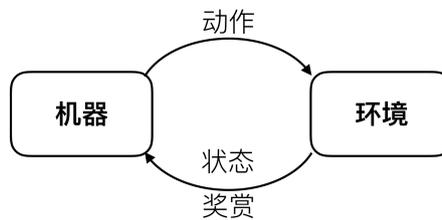


图 3.2 强化学习示意图

强化学习常见的应用有：下棋、机器人、自动驾驶等。在上文中提及的 AlphaGo 能够拥有强大的计算能力，模拟任一情况下的围棋步法，主要依赖于强化学习的算法。

除了以上三种机器学习方法，还有其方法的混合，比如说半监督学习 (Semi-Supervised Learning) 是原始数据部分有标签，部分没有。

监督学习中包含已知的输入和输出数据，可视为根据样本学习。无监督学习中没有输出数据，相当于自主学习，其学习目的是找到输入数据中存在的结构和模式。强化学习中既没有输入数据也没输出数据，只有特定的规则，相当于试错

学习。其目的是在多种可能路径中寻找最佳决策或者路线。在处理实际问题中，往往根据相应的情况选择其中一类方法或者几种方法的组合。

虽然机器学习学科的发展历史不过七八十年，但其发展迅速，且综合了数学、物理、生物、计算机等诸多学科的方法与技巧，内容繁多。机器学习方法主要分为概率性方法和非概率性方法，使用的模型主要有线性模型、神经网络、贝叶斯分类器、概率图模型等，采用的方法有决策树、支持向量机、集成学习、聚类分析、强化学习等。人工智能中几个主要概念的关系框架如图 3.3。

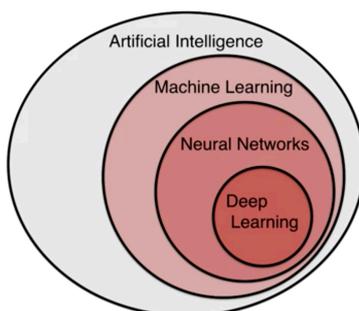


图 3.3 机器学习之间的相关关系

其框架简明扼要地说明了机器学习领域的脉络关系，人工智能是总领域，机器学习其主要方向，神经网络是机器学习的主要研究方法，深度学习是神经网络潜力巨大的发展对象。

3.2 机器学习的模型

由之前的关系图可知，神经网络是机器学习的主要模型。神经网络脉络复杂，具体方法实现贯穿在机器学习的整个过程中，以下主要介绍几个基本的神经网络结构。

3.2.1 神经元模型

神经网络（Neural Network）一词来源于生物学中的神经元（见图 3.4（a）），自然生物的大脑中神经元具有分辨、记忆、判别、计算等等功能，整合起来的杰出代表就是人类的智慧。据此，神经元的结构被其他各个学科所采用，尤其人工智能行业，据此发展出来的神经网络学在现今社会所指的默认为人工神经网络，而不是生物学意义上的神经网络，其在生物学神经网络的基础上改进颇多。神经网络是具有适应性的简单单元组成的广泛并行交互的网络，其组织能够模拟生物神经系统对真实世界物体所作出的交互反应^[8]。神经网络可以衍生出多种针

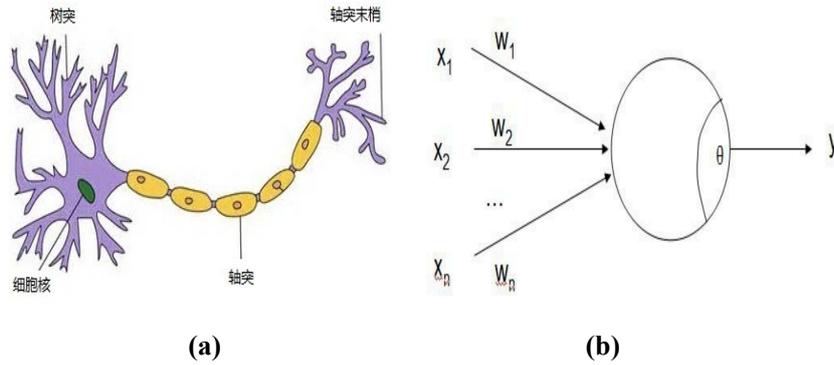
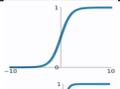
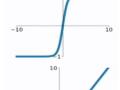


图 3.4 神经元模型和 M-P 神经元

为解决不同问题而设计的其他网络。

在生物网络中，每个神经元都与其他神经元相联，当其“兴奋”时，通过传递化学物质，改变神经元内的电位，若其超过一个“阈值”时神经元就会被激活，从而将化学物质传递下去。根据这一原理，“M-P 神经网络”得以出现^[9]。M-P 神经元模型示意图如图 3.4(b) 所示：在这个神经网络模型中，神经元接收到来自于 n 个其他神经元传递过来的输入信号，然后信号通过带权重的连接进行传递，神经元接收到的总输入值与神经元的阈值进行比较，然后通过“激活函数”来产生神经元的输出^[10]。在神经网络方法应用的大多实例中，激活函数的选取至关重要。常用的激活函数由：*sigmoid*, *ReLU*, *tanh* 等，激活函数没有最优的，只有对模型最适合的。常见的激活函数列举如下：

表 3.1 常用的激活函数

激活函数	表达函数	示意图
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	
tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	
ReLU	$f(x) = \max(0, x)$	

3.2.2 受限玻尔兹曼机

受限玻尔兹曼机 (**Restricted Boltzmann Machine**) 或者簧风琴是描述图模型如何应用于深度学习的典型模型^[10]。玻尔兹曼机是一种“基于能量的模型”，其神经元分为显层和隐层，显层用于表示数据的输入与输出，隐层则表示数据的内在表达。令向量 $s \in \{0, 1\}^n$ 表示 n 个神经元的状态， ω_{ij} 表示神经元 i 与 j 之间的

连接权, θ_i 表示神经元 i 的阈值, 则状态向量 s 所对应的玻尔兹曼机能量定义为:

$$E(s) = - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij} s_i s_j - \sum_{i=1}^n \theta_i s_i \quad (3.1)$$

若网络中的神经元以任意不依赖输入值的顺序进行更新, 则网络将达到玻尔兹曼分布, 此时状态向量 s 出现的概率将仅由其能量与所有可能的状态向量的能量有关:

$$P(s) = \frac{e^{-E(s)}}{\sum_t e^{-E(t)}} \quad (3.2)$$

玻尔兹曼机的训练过程就是将每个训练样本视为一个状态向量, 使其出现的概率尽可能大, 标准的玻尔兹曼机是一个全连接图 (见图 3.5)。

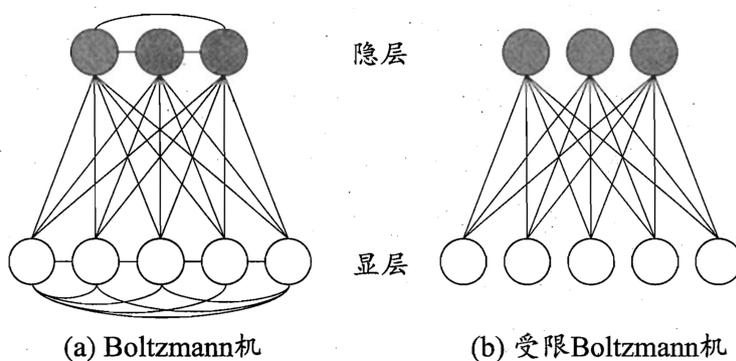


图 3.5 玻尔兹曼机示意图^[10]

标准玻尔兹曼机的训练的复杂度很高, 使其难以解决现实任务。实际上经常采用受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM), 其仅保留显层与隐层之间的连接。受限玻尔兹曼机常用“对比散度”(Contrastive Divergence, CD) 算法训练。假定网络中有 d 个显层神经元和 q 个隐层, 令 v 和 h 分别表示显层和隐层中的状态向量, 则由于同一层内不存在连接, 有:

$$P(v|h) = \prod_{i=1}^d P(v_i|h)P(h|v) = \prod_{j=1}^q P(h_j|v) \quad (3.3)$$

CD 算法对每个训练样本 v , 先根据上式算出隐层神经元状态的概率分布, 然后根据这个概率分布采样得到 h ; 此后, 类似地根据上式从 h 中产生 v' , 再从 v' 中产生 h' , 连接权的更新公式为:

$$\Delta\omega = \eta(vh^T - v'h'^T) \quad (3.4)$$

3.2.3 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是一种用来处理具有网格结构类型数据的神经网络，例如时间序列（在时间轴上采样形成的一维网格）和图像数据（二维的像素网格）。卷积网络是指那些至少在网络的一层中使用卷积运算来替代一般的矩阵乘法运算的神经网络。之所以采用卷积运算来构造模型，主要是卷积运算可以通过三个重要思想来改进机器学习系统：**稀疏交互、参数共享、等变表示**。

这三种思想使得在处理同一张图片时卷积方式比传统矩阵乘积的形式有效效率提高**40亿倍**。卷积网络一般分为三层：第一层并行计算多个卷积并产生线性激活响应；第二层每个线性激活响应通过一个非线性激活函数，称为探测级；第三层是池化层，使用某一位置相邻输出的总体统计特征来代替网络在该位置的输出^[11]。卷积网可以优化算法的复杂度，对具有时间序列的音视频处理效果良好，现代科技中图像识别、同声传译等技术等发展与此密切相关。图 3.6 是应用较广的手写数字识别的卷积神经网络图。

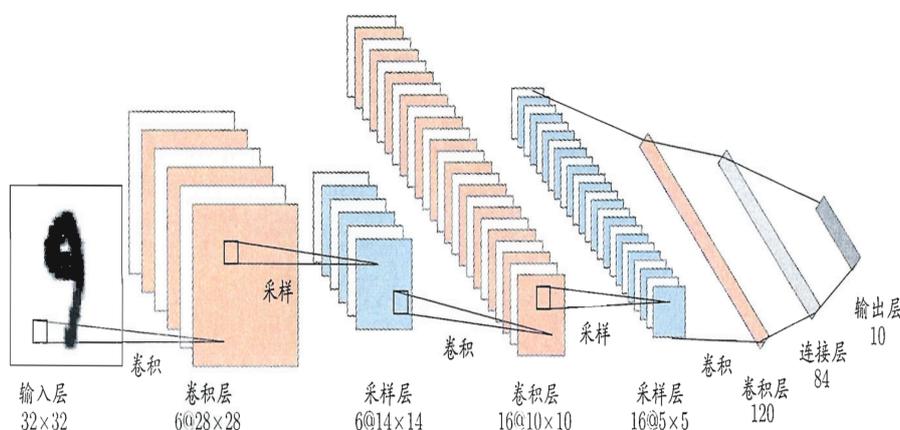


图 3.6 卷积神经网络用于数字识别的网络示意图

图 3.6 是著名的 LeNet-5 模型，是 LeCun 于 1998 年首先提出并成功应用于手写数字识别的神经网络模型^[12]，其对于机器学习界意义重大，对当前智能界起着里程碑的作用。

3.3 机器学习方法

主要介绍机器学习在物理中计算的几种方法。

3.3.1 PCA

主成分分析 (Principal Component Analysis, PCA) 是一种比较常用的降维方法, 可以将高维数据保留下重要的特征, 去除噪音和不重要特征, 其处理过程中会有一些的信息损失, 但可以节省大量的时间和成本。在利用 PCA 处理数据时, 首先将数据从原来的坐标系转换到新的坐标系, 由数据本身决定。转换坐标系时, 使用方差最大的方向作为坐标轴方向, 因为数据最大方差给出数据最重要的信息。第一坐标选择方差最大的数据, 第二组数据选取的是与第一组数据正交且数据方差次大的方向, 重复此过程, 坐标轴的方向数记为其特征维度^[10]。

PCA 算法的过程为:

Algorithm 1 PCA

- 1: 输入: 样本集 $D = x_1, x_2, \dots, x_m$;
 - 2: 低维空间维度: d' ;
 - 3: 过程:
 - 4: 对所有样本进行中心化: $x_i \leftarrow x_i - \frac{1}{m} \sum_{i=1}^m x_i$;
 - 5: 计算样本的协方差矩阵;
 - 6: 对协方差矩阵做特征值分解;
 - 7: 取最大的 d' 个特征值所对应的特征向量。
 - 8: 输出: 投影矩阵。
-

3.3.2 逆向误差传播算法

误差逆传播 (error BackPropagation, BP) 是在计算多层网络的杰出算法^[13], 也是至今最成功的神经网络算法。其训练过程是通过构建神经网络的权值, 求出其均方误差, 再基于梯度下降的策略, 以目标负梯度方向对参数进行调整, 求出均方误差最小情况下各个的权值与参数大小。其操作流程是: 先将输入样本提供给输入层神经元, 然后将信号逐层前传, 产生输出层的结果; 然后计算输出层的误差, 之后再将误差逆向传播至隐层神经元, 最后根据隐层神经元的误差来调整连接权和阈值, 循环其迭代过程, 直到达到停止条件。

其流程是:

Algorithm 2 BP

- 1: 输入: 训练集 $D = x_k, y_{k=1}^m$ 和学习率 η ;
 - 2: 过程:
 - 3: 在 $(0, 1)$ 范围内初始化模型中的权值和阈值;
 - 4: **repeat**
 - 5: **for all** $(x_k, y_k \in D)$ **do**
 - 6: 根据当前结构计算当前输出 \hat{y}_k ;
 - 7: 计算输出层和隐层的梯度项;
 - 8: 基于梯度下降策略, 以目标的负梯度方向调整, 更新模型的连接权与阈值。
 - 9: **end for**
 - 10: **until** 达到停止条件
 - 11: 输出: 连接权与阈值确定的多层前馈神经网络
-

3.3.3 核方法

在对训练样本分类的时候, 如果原始数据的维数是有限的, 则可以找到更高维上的高维特征使之可分。将 x 映射得到特征向量 $\phi(x)$, 在特征空间中划分超平面所对应的模型为^[7]: $f(x) = \omega^\top \phi(x) + b$, 其中 ω 和 b 是模型参量, 为最大化间隔, 有

$$\begin{aligned} \min_{\omega, b} \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i(\omega^\top \phi(x_i) + b) \geq 1, i = 1, 2, \dots, m \end{aligned} \quad (3.5)$$

转换为其对偶问题, 为:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \phi(x_i^\top) \phi(x_j) \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, m \end{aligned} \quad (3.6)$$

其中 $\phi(x_i)^\top \phi(x_j)$ 为 x_i 和 x_j 映射到特征空间的内积, 由于维数可能很高, 计算可能很复杂。因此设想函数: $\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \phi(x_i)^\top \phi(x_j)$ 即 x_i 和 x_j 在特征空间的内积等于它们在原始样本空间中通过函数 $\kappa(\cdot, \cdot)$ 计算的结果。重写其对偶问题为:

$$f(x) = \omega^\top \phi(x) + b = \sum_{i=1}^m \alpha_i y_i \phi(x_i)^\top \phi(x) + b = \sum_{i=1}^m \alpha_i y_i \kappa(x, x_i) + b \quad (3.7)$$

这里的 $\kappa(\cdot, \cdot)$ 函数就是核函数, 由上式可知其模型最优解可通过训练样本的核函数展开, 称作“支持向量展式”。核函数的定义是: 令 $\kappa(\cdot, \cdot)$ 为输入空间, $\kappa(\cdot, \cdot)$ 为定义在 $\mathcal{X} \times \mathcal{X}$ 上的对称函数, 则 κ 是核函数当且仅当对于任意数据

$D = \{x_1, x_2, \dots, x_m\}$, “核矩阵” (kernel matrix) K 总是半正定的。

其表明, 只要一个对称所对应的核矩阵半正定, 其就能作为核函数使用。如果希望样本在特征空间内线性可分, 则特征空间的好坏对支持向量的性能至关重要。“核函数的选择”对支持向量结果的好坏影响重大。常用的核函数如表 3.2 所示。

表 3.2 常用的核函数

名称	表达式	参数
线性核	$\kappa(x_i, x_j) = x_i^T x_j$	
多项式核	$\kappa(x_i, x_j) = \{x_i^T x_j\}^d$	$d \geq 1$ 为多项式的次数
高斯核	$\kappa(x_i, x_j) = \exp(-\frac{ x_i - x_j ^2}{2\sigma^2})$	σ 为高斯核的带宽
拉普拉斯核	$\kappa(x_i, x_j) = \exp(-\frac{ x_i - x_j }{\sigma})$	$\sigma > 0$
Sigmoid 核	$\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$	\tanh 为双曲正切函数

此外, 还可以通过核函数的组合得到新的核函数。

机器学习与物理通过交叉来对物理相关材料、属性进行分析计算的过程需要综合两方面的知识, 追本溯源, 因此, 其采用的方法往往不是单纯的一种, 而是根据不同的物理问题, 具体问题具体分析, 综合几种不同的机器学习方法, 相互独立或者结合求出相关问题的解。

第4章 机器学习在强关联体系中的应用

强关联电子体系在凝聚态物理的发展中有着举足轻重的地位，也是当今世界的物质研究前沿问题。在相关量子理论逐渐完善的基础上，强关联电子体系的计算问题在遭遇指数墙困难的情况下，结合现今流行的机器学习理论与其处理复杂数据的能力，相互渗透，相互交叉，使其能够相互合作得出令人满意的结果。以下为将机器学习应用到物理中的几个例子，借此说明其交叉的优势。

4.1 对相的分类

现代机器学习结构，比如全连接卷积神经网络，能够在各种凝聚态物理哈密顿量中辨别相与相变。通过现代软件库的简单编程，直接从初始态利用蒙特卡洛特征采样，可以训练神经网络探别多种序参量和非传统序的高度非平凡态。

物质的阶段划分和识别它们之间的相变是凝聚态物理研究的核心话题之一。尽管构成粒子的数量是天文数字，但它往往足以代表只有少数变量的多体系统的状态。例如，传统的物理学方法是通过对称性考虑或者分析低能量集体自由度来识别有序参数，并用它们来标记物质的相位。然而，在越来越多的物质新状态中，以这种方式识别阶段和相变是比较困难的，其中阶次参数只能以难以捉摸的非局部方式来定义。这些新的发展需要寻找适当的相变指标的新方法。

1. 背景介绍

一、相变以及 Ising 模型

相变是物理学界充满意外发现与未知的一个领域，其现象丰富多彩，可以从不同的角度分类和研究。一般来说，相变是有序和无序相互竞争的结果，有序的起因是相互作用，无序的来由是热运动，伴随着有序无序过程的转变是体系对称性的变化^[14]。相变点附近变化的参数称为序参量在凝聚态物理中，序参量是一个描述体系发生奇异现象的重要参数。为研究体系的相变随序参量的变化，引入机器学习方法进行研究。可以利用 Ising 模型进行探究。

Ising 模型是一个经典的铁磁—顺磁相变模型，是一个单轴的离散自旋模型，即其自旋取值只能沿单轴取 +1 或者 -1 值，其允许临近自旋的相互作用，是描述相变的经典模型。Ising 模型的序参量是其磁矩，随着温度在相变点附近的改变，其磁矩改变，即高温为顺磁项，低温变为铁磁相。本例中采用二维 Ising 模型描述相的变化，根据二维 Ising 模型的昂萨格解其理论上的相变为： $T/J = 2.269$ ^[15]，

其哈密顿量为：

$$H = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z \quad (4.1)$$

设 $J=1$, $\sigma_i^z = \pm 1$, 所以对于 N 个格点, 其态空间是 2^N 的大小。

二、MNIST 原理

MNIST (Mixed National Institute of Standards and Technology) 是一个大型的手写体数字识别数据库, 在计算机领域中利用这个库和 TensorFlow 框架能够实现基本的阿拉伯数字识别, 即: 给定一个手写数字, 计算机能够在训练之后自己判断其数字含义。在实际的实现中, 除了给定的数字识别, 经过发展还可以识别其他的内容, 比如说汉字或者其他语言文字等。

手写数字识别的过程一般采用: 输入图片—提取特征—分类器—输出结果的流程。其中关键点是特征提取的过程。特征提取一般采用 LeNet-5 的模型, 即图 3.6 介绍的卷积网络。卷积神经网络提取图片特征的原理第一步是利用一个核来定义权值矩阵, 然后将其矩阵在图片上进行扫描, 相当于权值共享, 经过这一步之后相当于将图片进行了滤波, 采用特定的权值矩阵能够提取图片中的边缘信息或者特定颜色, 其中权值是通过训练学得的。这一步操作相当于是卷积网络中的卷积层操作, 之后经过为降低网络维度而处理的步骤称为池化层和最后的分类输出层, 卷积操作的以训练完成。其中搭载了较为复杂的数学知识和网络构架, 因此一般采用现有的 Google 已经实现的网络框架进行实现。

在本文中, 为了实现物理中相变的自动识别, 利用了以上的手写识别数字的原理^[16]。

2. 模型以及数据处理

本例仿真采用了 python3.6.4 版本, 引用了 TensorFlow 框架, 借鉴了 YanLe-Cun 关于卷积神经网络的构想, 参考了 Google 的 MNIST 的构造, 来仿真求解二维方格子 Ising 模型的情况。

(a) 产生样本

机器学习的训练需要大量的样本, 对于本例采用的监督学习算法的模拟, 需要利用模型相变前后的的大量数据以及其标签。利用计算模拟中最强大的工具: 蒙特卡罗模拟来构造样本集。

二维 Ising 模型中粒子的分布采用玻尔兹曼分布, 即其分布与其温度有关: $f(T) \propto \exp(-\frac{E}{kT})$, 在利用蒙特卡洛方法构造样本时, 要达到所要求的样本与其温度相关, 需要其满足细致平衡条件, 即在满足给定温度下随着每一次迭代能量降低的前提下, 还有一定概率接受相反的情况, 在足够次迭代之后能够使其数据

集收敛到给定的分布。流程图如下（采用 28×28 的二维周期性格点）：

Algorithm 3 蒙特卡洛产生样本算法

- 1: 输入：初始随机产生的样本：784 个随机自旋上下的格子；
 - 2: 流程：
 - 3: **repeat**
 - 4: **for** 当前每一个状态 **do**
 - 5: 1、利用其哈密顿量求解当前格点的能量；
 - 6: 2、随机选取一个格点翻转；
 - 7: 3、计算翻转后与翻转前的能量差。
 - 8: **if** 翻转之后能量降低 **then**
 - 9: 接受翻转之后的格点；
 - 10: **else** 翻转之后能量增多
 - 11: 产生一个 $(0, 1)$ 范围内的随机数 m ；
 - 12: **if** $m \leq \exp(-\frac{E}{kT})$ **then**
 - 13: 接受翻转；
 - 14: **end if**
 - 15: **end if**
 - 16: **end for**
 - 17: **until** 达到足够的次数，格点几乎不再翻转。
 - 18: 输出：格点的状态。
-

实验生成了 $1.0 \sim 5.0$ 温度范围，每隔 0.1 的间隔取 1000 个样本，并加上合适的标签，生成图的形式如图 4.1 所示。

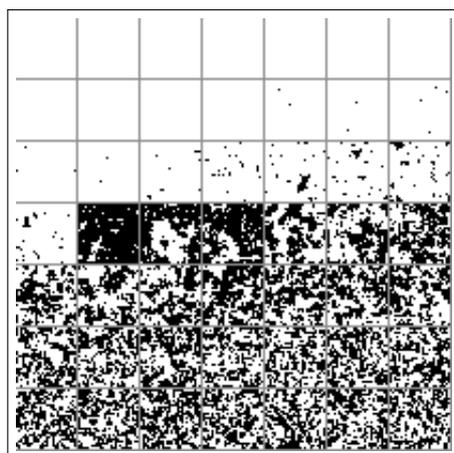


图 4.1 $0.1 \sim 5.0$ 的二维 Ising 模型，白色代表自旋向上，黑色为自旋向下

这些数据是未经处理的大量原始数据，在之后的网络中能够直接采用，不需要进行预处理的过程。

(b) 构造模型

采用全连接前馈神经网络的模型（前馈是指无环或者回路），分为三层：输入层，隐层，输出层。输入层由自旋特征决定，隐藏层是 100 个 sigmoid 神经元，输出层是分类的特征。示意图如下：

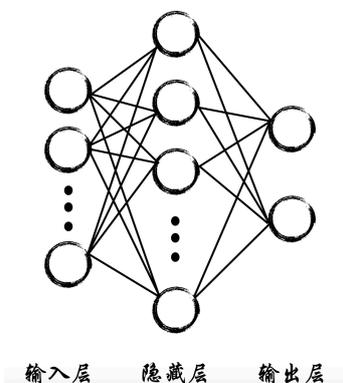
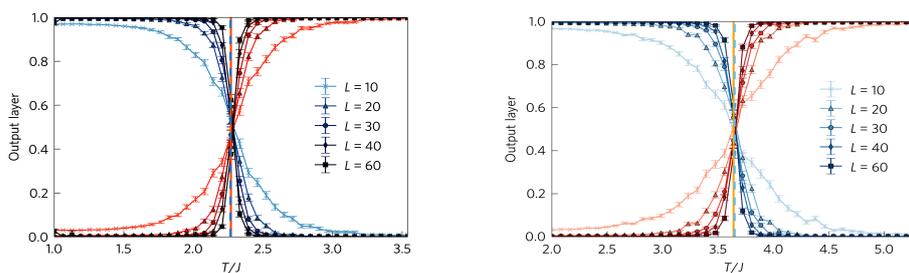


图 4.2 神经网络

(c) 数据处理与分析

经由以上训练产生的模型是其自动根据输入的图片判别其为铁磁相或者反铁磁相的过程，与样本输入的结构、温度等其他因素无关。由于本人能力所限，神经网络没有构建成功，为介绍此种方法的结果，引用参考 Juan Carrasquilla 和 Roger G.Melko 的文献^[16]的结果，如图 4.6 所示。其中 4.6 (a) 中训练方格子的相变点为： 2.266 ± 0.002 ，与理论计算结果相同。并且，作为验证，将具有三角格子结构的 Ising 模型图片作为输入层输入已经训练完成的全连接前馈神经网络，可以得到图 4.3(b)。



(a) 训练方格子模型得到的神经网络

(b) 已得网络求解三角格子的结果

图 4.3 神经网络输出层^[16]

据图 4.3(b) 可知，其相变温度点为： 3.65 ± 0.01 ，与理论计算相符合。说明其方法的可行性以及推广性。

3. 结果与分析

以上根据已加标签的训练样本进行分类，对于同一类型的新样本，可以利用

其已训练好的模型进行分类。以上是根据监督学习据给定的训练集的标签经过训练而得到的，而在有些情况下，其相并不能确定标签，此时可以采用无监督学习的方法，无监督学习是从数据中寻得结构，采用聚类和 PCA 方法求解样本中的结构，从而对其进行分类^[17]。机器学习方法引入的目的是：期待能够在未知能量或者哈密顿局域条件未知的情况下能够从相的序元素中学得有序相。机器学习能够在具有不同特征的样本下进行根据其特征进行分类，同时，也有潜力表示量子物理中的波函数。

4.2 研究波函数与基态能量

量子物理中多体问题的复杂性来源于其多体波函数之间的指数性的关联增长。但是通过构建波函数系统的机器学习网络就能将其转变为计算机能够解决的问题。经演示的强化学习方案既能找到基态，又能描述复杂相互作用且单一时间演化的量子系统。尤其在描述一维或者二维的自旋相互作用模型时，非常精确。

1. 波函数

波函数 ψ 是量子物理中的函数概念在难以在经典世界中表述，其在量子态中代表一个重要的数学量，不仅仅能够单粒子态，也能够表示复杂的分子态。全编码所有的物理态需要指数形式的信息量，但波函数表示的话可以使用比相应希尔伯特空间小的量代表。利用波函数形式可以使用有限的资源来解决多体薛定谔问题。由于其方法是利用神经网络求解量子态，因此简称为 NQS (Neural-network Quantum States)^[18]。

2. 模型

考虑一个有个 N 个离散值的量子系统，其自由度为： $S = (S_1, S_2, \dots, S_N)$ ，其基本粒子可能是自旋，玻色占据数或之类的。机器学习的思路是：将波函数看做是一个黑箱，输入是多体参数 S ，输出是根据 $\psi(S)$ 得出的振幅与相位，目的是将作为一个黑箱的神经网络训练成波函数 $\psi(S)$ 。不同的网络结构用来解决不一样的特定问题，对于多体来说，最好的结构也不是唯一的。为了具体化，利用玻尔兹曼机 (RBM) 来描述自旋为 $1/2$ 的量子系统，在这种情况下，玻尔兹曼机由一个与选定基矢 (比如说： $S = \sigma_1^z, \dots, \sigma_2^z$) 的物理自旋变量对应的 N 节点可视层，有 M 个辅助自旋变量 (h_1, \dots, h_M) 的单隐层组成^[18]。这个描述对应着量子

态的多变量表示:

$$\Psi_M(S; W) = \sum_{\{h_i\}} e^{j \sum a_j \sigma_i^z + \sum_i b_i h_j + \sum_{ij} W_{ij} h_i \sigma_j^z} \quad (4.2)$$

式中 $h_i = \{-1, 1\}$ 代表一组 M 个隐藏的自旋变量, 神经网络元素 $W = \{a, b, W\}$ 代表给定的输入态 S 。因为连接层之间没有相互作用, 因此隐藏参数能够被完全描绘出来, 波函数读作: $\Psi(S; W) = e^{j \sum a_j \sigma_j^z \times \prod_{i=1}^M F_i(S)}$, 其中 $F_i(S) = 2 \cosh[b_i + \sum_j w_{ij} \sigma_j^z]$ 其中的权重完全描述整个波函数的振幅和相位。

NQS 的数学基础建立在其能够形容复杂的多体波函数的可靠理论上, 保证有神经网络近似等价于高维平滑规则的函数。如果这些条件被多体波函数满足的话, 就可以证明 NQS 方法成立。这种方法的优点在于其性能能够通过系统增加隐层的数目而得以提升。

具体来说, 对于整数隐藏变量密度 $\alpha = 1, 2, \dots$, 权值矩阵采用特征采取器 $W_j^{(f)}$ 其中 $f \in [1, \alpha]$ 。这些挑选器由 αN 的元素来代替之前非均匀下的 αN^2 的变量。剩下的任务就是优化网络中的参数 W 。在量子力学中, 多体的状态是未知的, 即没有确定的样本, 因此, 监督学习不能够解决这一诉求。相反, 我们得到了一种一致的强化学习方法, 在这种方法中, 基态波函数或依赖于时间的函数是根据变分原理的反馈而学习的。

为了证明 NQS 在形容复杂多体运动中的准确性, 增强学习通过最小化基态未知的能量 $E(W) = \langle \Psi_M | H | \Psi_M \rangle / \langle \Psi_M | \Psi_M \rangle$, 每次迭代通过梯度优化逐步向能量最低的方向进行, 最终达到最低。

3. 分析

为了说明方案, 利用两个典型自旋模型的基态的问题: 转场-场化 Ising (TFI) 模型和反铁磁海森堡 (AFH) 模型。他们的哈密顿量是:

$$H_{TFI} = -h \sum_i \sigma_i^x - \sum_{ij} \sigma_i^z \sigma_j^z \quad (4.3)$$

和

$$H_{AFH} = \sum_{ij} \sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z \quad (4.4)$$

其中, $\sigma^x, \sigma^y, \sigma^z$ 是泡利矩阵。

考了一维和二维的周期性边界条件格点, 网络能够根据其模式从基态波函数中提取特定的关联参数, 对于二维的格点来说, 其对应着反铁磁性关联。除此之外, 网络模型还能学得模型的基态能量, 经过更进一步的改进, 也能求解含时间问题的薛定谔方程。

据此，基于神经网络的方法可以被成功地用来解决量子多体问题，不仅是为了基态的特性，还可以用来模拟由一组复杂的激量子态引起的进化。人工神经网络在变分量子态中能够有效地计算一维、二维的纠缠多体系统。尽管受限玻尔兹曼机在这里只被简单地应用，却求出了经典的自旋系统的基态和动力学属性。能够期待在更进一步的发展中，量子力学的更多结构能被挖掘出来。

第 5 章 当前进展

现代机器学习研究的中心目标是直接从数据中学习和提取重要特征，物理方法探求的是从实际现象中找寻其根本规律与代表的特征，两者殊途同归。鉴于机器学习近年来取得诸多进步，其方法和系统逐步完善，成熟，物理学方面开始逐步引入其方法来促进物理方面的进步以及新的突破。

目前，机器学习部分网络中的学习体系已经逐步趋于完善，包括监督学习、无监督学习、强化学习的网络在物理上的应用，目前已经实现经由监督学习、无监督学习分析得出传统模型的相变，拓扑序；针对量子物理方面，经由设计得到的神经网络，能够求的模型的波函数以及基态能量，利用机器学习中内核方法的内涵，解决量子计算的一些问题；针对物理方面的一些方法，机器学习中的深度学习可以与 RG（重整化群）相对应，神经网络可以与 DMRG（密度矩阵重整化）相对应，卷积网络可以与 TensorNetwork（张量网络）对应，玻尔兹曼机可以与 BZ（玻尔兹曼函数）对应等等^[19-22]。此外，也可以将物理上面的一些概念，比如说张量、量子态、热力学分布应用到机器学习上，促进其发展与理解。由此可见，机器学习与物理两方的交流，能够同时促进双方领域的发展与进步。

将机器学习方法应用到物理方面，除了其可见的成效之外，还有一些不可不忽视的问题。机器学习经过将近八十年的发展，期间几次衰败，几次又兴盛起来，期间离不开机器学习深度学习三巨头：Hinton、LeCun、Bengio 的不断发展以及新思路的注入，尤其是 2016 年 AlphaGo 打败李世石，震惊世界。其卓越的实用性以及成效显著的功能能够使其如今发展如火如荼，但是其理论基石涉及到复杂至极的数学理论，很多模型训练的特征属性尚不能完全把握。此外，很多物理学处理的问题大都是根据建立的模型来研究的，其内部的对称性以及各种条件限制如何更好地嫁接到机器学习方法中此前也是值得考虑的问题。

总而言之，机器学习在强关联系统中跨过指数墙困难，直接求解其问题，这一点可以带动物理诸多方面的发展。如何在当前的情景下，趋利避害，有效结合两个的优势，发展学科的进步，是当前物理人和计算机人的共同诉求。

第6章 未来展望

神经网络和机器学习经过几十年的改进，现在已经进入了一个新时代，其应用将在许多科学，工业和技术领域爆炸性发展。机器学习源于对人工智能的追求，是当今跨学科领域最活跃的领域之一，具有从化学信息学，生物学，宇宙学等基础研究到量化社会科学的广泛应用。在物理学中，最近引入了引力波分析，黑洞检测和材料设计等机器学习技术^[23-25]。

机器学习在物理方面有很多值得推广的应用，比如说：回归方法用于预测晶体结构，近似密度函数，以及解决量子杂质问题，训练神经网络来分类经典统计模型^[17]。未来其发展的方向除了机器学习中需要更加强大理论的支撑之外^[26]，机器学习中其他的先进算法如深度学习、卷积网络、深度学习、胶囊网络等也需要以不同的方式来引入到物理中，来进行物理中固体、量子物理、热力学、量子计算机等各个方面的更深层次的发展^[27]。

机器学习方法的引入不仅仅能够从新的角度来思考物理，解决物理问题，而且由于机器学习方法的自动智能性，期盼人类能够创造出一个胜于人脑智慧机器，以期能够解放人力，促进人类社会的跨越式进步。

参考文献

- [1] 蔡子. 机器学习方法在量子多体物理中的应用[J]. 物理, 2017, 46(9): 590–596.
- [2] COLEMAN P. Introduction to many-body physics[M]. [S.l.]: Cambridge University Press, 2015.
- [3] R BULLA A C H, PRUSCHKE T. Numerical renormalization group calculations for the self-energy of the impurity anderson model[J]. Science, 1998.
- [4] NOACK R M, WHITE S R. Density-matrix renormalization: A new numerical method in physics[M]. Berlin Heidelberg: Springer, 1999.
- [5] 黄理. 强关联电子体系的 LDA+DMFT 第一性原理计算方法[D]. 北京: 中国科学院物理研究所, 2012.
- [6] NENG XIE D H Y F Y. Hybridization oscillation in the one-dimensional kondo-heisenberg model with kondo holes[J]. Scientific Reports, 2017.
- [7] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- [8] KOHONEN. An introduction to neural computing[J]. Neural Networks, 1988, 1(1): 3–16.
- [9] SuPhoebe. M-p 模型的来源[Z]. <https://blog.csdn.net/u013007900/article/details/50066315>, 2015.
- [10] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [11] YANN LECUN Y B G H. Deep learning[J]. nature, 2015, 521: 436–444.
- [12] LECUN Y B P, L.Bottou. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998.
- [13] YOSHUA BENGIO A C. Deep learning[M]. 赵申剑, 李凯等. 北京: 人民邮电出版社, 2017.
- [14] 于渌. 边缘奇迹相变和临界现象[M]. 北京: 科学出版社, 2005.
- [15] KRAUTH W. Statistical mechanics algorithms and computations[M]. Paris: Oxford University Press, 2005.
- [16] CARRASQUILLA J, G.MELKO R. Machine learning phases of matter[J]. nature physics, 2017, 17(6): 413–415.
- [17] Lei Wang. Discovering phase transitions with unsupervised learning[J]. Phys. Rev. B, 2016. DOI: [10.1103/PhysRevB.94.195105](https://doi.org/10.1103/PhysRevB.94.195105).
- [18] CARLEO G, TROYER M. Solving the quantum many-body problem with artificial neural networks[J]. Science, 2017.11.
- [19] LEVINE Y, YAKIRA D, COHEN N, et al. Deep learning and quantum entanglement: Fundamental connections with implications to network design[J]. CoRR, abs/1704.01552, 2017.
- [20] HUGGINS W, PATEL P, WHALEY K B, et al. Towards quantum machine learning with tensor networks[J]. arXiv preprint arXiv:1803.11537, 2018., 2018.

- [21] STOUDEMIRE E M, SCHWAB D J. Supervised learning with quantum-inspired tensor networks[J]. Journal of Statistical Physics, 2017.
- [22] KOCH-JANUSZ M R Z. Mutual information, neural networks and the renormalization group [J]. Nature Physics, 2018.1.
- [23] J B. Quantum machine learning matrix product states[J]. arXiv preprint arXiv:1804.02398, 2018.
- [24] MEHTA P S D J. An exact mapping between the variational renormalization group and deep learning[J]. arXiv preprint arXiv:1410.3831, 2014.
- [25] DENG D L S S D, Li X. Machine learning topological states[J]. Physical Review B, 2017.
- [26] LI L W S R E A, Baker T E. . pure density functional for strong correlation and the thermodynamic limit from machine learning[J]. Physical Review B, 2016.
- [27] SCHULD M K N. Quantum machine learning in feature hilbert spaces[J]. arXiv preprint arXiv:1803.07128, 2018.

致 谢

感谢在毕设期间在物理所给予我学术上帮助的杨义峰老师，胡丹青师兄、黄东宸师兄，还有徐远骥师兄的 wifi，李宇师兄的演讲、韩儒磊师兄的报销单，董建军的兔子肉、王倩倩师姐的各种零食、刘哲师兄的各种传奇事件，当然还有同期的邹博同学的各种照片段子，武振超同学的帮助，蔚紫冰与邹雨婷舍友的相互吐槽。此外，也感谢中南大学的老师易有根易大师的校内支持，没有你们，我的毕设生活不会如此丰富多彩。大学四年转眼即逝，一切以毕设为笔尾将要画上一个句号。在这四年中，我学会了珍惜，学会了成长，学会了责任，在毕业之际，除了对将来研究生博士生涯的期待之外，还有对长沙生活的不舍，对中南的眷恋与归属感，对在这四年认识的朋友们的离愁，这四年的欢声笑语，烦愁别绪，终于洒在每个成长在这片土地上的我们身上，我们将承载着这一份记忆，继续前行。

莎士比亚说过一句话：凡是过往，皆为序章 (What's past is prologue)，以往的只能算是一个引子，不论是灰暗，还是荣耀，那些曾经竭尽全力努力过的都已经过去，希望在未来的研究生生涯、博士生涯中，新的篇章的开启。

致敬，感谢，前行！

