

前言

由于深度学习有从复杂系统中提取基本特征的强大能力，它在图像识别，语音处理，生物系统等领域都具有广泛的应用。特别是在凝聚态物理领域，深度学习展现了可以精确逼近复杂物理系统的独特能力。在过去的几年，传统的神经网络在凝聚态物理领域无论是在提取特征还是在生成模型方面都取得了不错的成果。然而由于构型尺寸的限制，导致神经网络只能生成单一尺寸的构型，限制了神经网络的泛化性而且传统的深度学习算法假设数据样本之间彼此独立，但是在实际的物理构型中，每个自旋节点都会与构型中的其他节点以及边相关。图神经网络的出现解决了这一问题，图是不规则的，它可以适应任意尺寸的构型，并且可以捕获不同节点之间的相互依赖关系。变分自编码是一种无监督学习算法，可以有效地模拟训练数据的分布。在本文中，我们使用图神经网络这一特性与变分自编码相结合，构建Ising模型模拟器，生成的Ising模型与传统蒙特卡罗方法模拟的数据非常相似。结果表明，图神经网络在凝聚态物理领域是一种有潜力的计算工具。

引言

今年来，物理学家已经开始采用深度学习技术来研究凝聚态物理领域。大多数任务都是通过监督学习算法来完成的。在监督学习中，算法是对带标签的数据进行训练，将标签分配给数据点。经过成功的训练，可以对以前未见过的数据的标签进行高精度的预测。除了有监督学习以外，还有一些无监督学习算法可以在无标签数据中找到结构。已经有可能使用无监督的学习技术来重现Ising模型的蒙特卡罗采样状态，以无监督的方式发现相变已经成为主流。受限制玻尔兹曼机，变分自编码和生成对抗网络等都在Ising模型的模拟上广泛应用，并取得了不错的效果。但是这些模型在训练时都无法兼顾物理模型分子与分子之间的相互作用和分子之间边的影响。同时很多物理模型的微观结构时不规则的，生成模型的泛用性差，一般只能适应一种温度下的一种尺寸。这就需要一种能考虑非结构化数据，并且能考虑节点与节点，节点与边的关系的深度神经网络。

幸运的是，近年来深度学习领域关于图神经网络的研究热情日益高涨，图神经网络处理非结构化数据的出色能力使得其在物理，生物和化学等方面的应用尤为突出。大多数物理模型本质上就是一种图结构。图卷积网络可以聚合邻居节点的信息，相比较传统卷积网络，图卷积网络更善于编码图的结构信息，能够学习到更好更深层次的表示特征。图变分自编码是一种无监督图生成算法，是图神经网络与生成网络相结合的产物，它能够有效利用数据的图结构去模拟训练数据的分布。

在这篇文章中，我们将图卷积网络整合到图变分自编码器框架中构建Ising模型模拟器，用来模拟不同温度下的Ising模型构型。经过训练的模拟器能够有效的生成Ising模型状态，其物理特性与传统蒙特卡罗模拟得到的结果并无不同。并且可以适应多种不同的尺寸并且极大的缩短了模拟时间。我们将首先介绍一般形式的Ising模型，图变分神经网络与图卷积网络。在文章的核心部分，我们构建一个与图卷积网络相结合的图变分自编码器，模拟不同温度下的2D Ising模型，并通过将生成的数据与来自蒙特卡罗模拟的数据进行比较来验证我们方法的有效性模拟。

Ising模型与图神经网络

Ising模型

在统计物理中, Ising模型被描述为某些晶格中具有耦合相互作用的二元自旋的集合。考虑N个自旋 $s=\{s_i\}$ 可以取值 ± 1 ,其中索引标记自旋 s_i 的位置,Ising系统的标准哈密顿量仅包括最近邻相互作用,并且每个自旋方向可能是"向上"(+1)或"向下"(-1),尽管广义模型可能包括长程相互作用和更多的自旋选择方向.标准哈密顿量是:

$$H = -J \sum_{neighbors} S_i * S_j$$

图神经网络

图是一种数据结构,它对一组对象(节点)及其关系(边)进行建模。今年来,由于图结构的强大表现力,用机器学习方法分析图的研究越来越受到重视。图神经网络是指使用神经网络学习图结构数据,提取和发掘图结构数据中的特征和模式,满足聚类,分类,预测,分割,生成等图学习任务需求的算法总称。图神经网络发展的一个动机源于卷积神经网络。卷积神经网络的广泛应用带来了机器学习领域的突破并开启了深度学习时代。然而卷积神经网络只能在规则的欧式空间数据中提取潜在特征,不能很好的对应现实中复杂多变的图数据,如何将卷积神经网络应用于图结构这一非欧式空间成为图神经网络模型重点解决的问题。以下给出图神经网络中相关符号的说明和定义

定义1. 图是由节点和连接节点的边所构成的,通常记为 $G = (V, E)$ 。其中 $V = \{v_1, v_2, \dots, v_n\}$ 代表节点集合, $E = \{e_1, e_2, \dots, e_n\}$ 代表边集合,边也可以用 (v_1, v_2) 的方式来表示.通常节点也被称为顶点或者交点.边也被成为链接或者弧.通用的图表示是一个五元组: $G(V, E, A, X, D)$.其中A代表图的邻接矩阵,X代表节点的特征矩阵,D代表度矩阵.

定义2. 邻接矩阵:图的邻接矩阵指用于表示图中节点的连接情况的矩阵.该矩阵可以是二值的,也可以是带权的.对于有N个节点的无向图来说,邻接矩阵是一个N*N的实对称矩阵.

定义3. 度矩阵:节点的度表示与该节点相连的边的数量.图的度矩阵即用于描述图中每个节点的度的矩阵.度矩阵是一个对角矩阵,对于无向图来说,一般只使用入度矩阵或者出度矩阵.

定义4. 组合拉普拉斯矩阵,又称标准拉普拉斯矩阵,由对角矩阵和邻接矩阵组合而成:

$$L = D - A$$

该矩阵只在中心节点和一阶相连的节点上有非零元素,其余之处均为零.拉普拉斯矩阵也是图的一种表现形式. **定义5.** 归一化拉普拉斯矩阵

$$L^{sym} = I - D^{-1/2} A D^{-1/2}$$

其元素值为:

$$L_{(i,j)}^{sym} = \begin{cases} 1, & i = j \text{ 且 } deg(v_i) \neq 0 \\ \frac{-1}{\sqrt{deg(v_i)deg(v_j)}}, & i \neq j \text{ 并且 } v_i \text{ 与 } v_j \text{ 相连} \\ 0, & \text{otherwise} \end{cases}$$

其中 $deg(v)$ 表示为节点 v 的度

图卷积网络

图卷积网络是由传统的卷积神经网络引申出的图卷积网络，图卷积可以两种，基于频谱的方法和基于空间的方法。基于频谱的方法从图信号处理的角度，引入滤波器来定义图卷积，因此基于频谱的图卷积可以理解为从图信号中去除噪音。基于空间的图卷积方法通过汇集邻居节点的信息来构建图卷积。

在了解图卷积之前，我们先要对卷积操作有一个基本的概念。卷积网络最开始应用在图像上，我们将图像上的像素点作为特征来进行卷积操作提取特征。用随机的共享的卷积核得到像素点的加权和从而提取到某种特定的特征，然后用反向传播来优化卷积核参数就可以自动的提取特征，是卷积神经网络提取特征的基石。可以这么认为，卷积操作相当于将周围的像素点信息聚合到卷积核中央像素点上，提取出该部分的特征。

所以我们在图领域想利用图像领域的卷积，使用一个通用的范式来进行图特征的提取，与图像类似，我们在想得到节点的特征表示时，可以通过聚合邻居节点的信息，可以这样理解，图中的每个节点无时无刻不因为邻居节点核更远的点影响而在改变自己的状态直到最终的平衡，关系越亲近的邻居影响越大。

图卷积通式

对于图卷积网络来说，我们想要学习到它对于每个节点的特征表示，任意的图卷积都可以写成这样的一个非线性函数：

$$H^{l+1} = f(H^l, A)$$

$H^0 = X$ 为第一层的输入, l 代表神经网络的层数, A 代表邻接矩阵.

传统的图卷积网络做法就是将邻居节点的信息加入到本节点当中:

$$f(H^l, A) = \sigma(AH^lW^l)$$

W^l 代表第 l 层的参数矩阵, σ 代表激活函数. 根据矩阵乘法可以看出上式中的每个节点都结合了相邻节点的信息,但由于仅仅使用了相邻矩阵,相邻矩阵对角线为0,无法体现节点自身的信息,所以我们使用拉普拉斯矩阵来代替邻接矩阵

$$f(H^l, A) = \sigma(LH^lW^l)$$

上式中引入了拉普拉斯矩阵,从而解决了没有考虑自身节点信息自传递的问题,但是由于没有被规范化,我们将拉普拉斯矩阵规则化得到图卷积

$$f(H^l, A) = \sigma(L^{sym} H^l W^l)$$

上面讲述的都是以矩阵形式计算,我们也可以从单个节点角度来观察公式,对于第 $l + 1$ 层的节点特征 h_i^{l+1} ,对于他的邻接结点 $j \in N$, N 是结点 i 的所用邻居结点的集合,所以图神经网络的更新公式同样可以描写为

$$h_i^{l+1} = f(h_i^l) = \sigma(h_i^l \cdot W_1^{l+1} + \sum_j h_j^l \cdot W_2^{l+1})$$

变分自编码器

变分自编码器是一种变分推理的生成模型,由编码器和解码器两部分网络组成,它是包含隐变量的一种模型.在变分自编码器当中,我们基于这样的假设: 我们的样本 x 是某些隐变量 z (latent variable)通过某种映射产生出的,而 z 也不是一个固定值,而是服从一个分布: $z \sim P_\theta(z)$,则 $x \sim p_\theta(x|z)$,这里的 P 是由参数 θ 决定的分布族,而 θ 就是我们需要找到的真实分布.所以我们最根本的目标就是确定 z 和 θ 的值,这样就能够得到样本 x .为了推断 $p_\theta(x|z)$,我们需要最大化边际对数似然 $\log(p_\theta(x))$ 我们可以重写该式:

$$\begin{aligned} \log(p_\theta(x)) &= \log \int p_\theta(x|z) dz = \log \int p_\theta(x|z) p(z) dz \\ &= \log \int p_\theta(x|z) p(z) \frac{q_\phi(z|x)}{q_\phi(z|x)} dz \\ &\geq \mathbb{E}_{z \sim q_\phi(z|x)} [\log(p_\theta(x|z))] - D_{KL}[q_\phi(z|x) || p(z)] \\ &:= \varepsilon(x, \theta, \phi) \end{aligned}$$

我们在第四步应用了Jensen不等式.在第二步中对潜在变量 z 的积分通常是棘手的,因此我们引入了带有参数集 ϕ 的近似后验分布 $q_\phi(z|x)$ 并使用变分推理原理来获得边缘对数似然的易处理界限,即证据下界(ELBO).我们使用 $p(z) = N(1, 0)$ 作为潜在变量先验.ELBO是对数似然的易处理下界,因此可以最大化推断 $p_\theta(x|z)$. $\mathbb{E}_{z \sim q_\phi(z|x)} [\log(p_\theta(x|z))]$ 可以理解为重构误差,因为最大化它会使解码器的输出类似于编码器的输入, $D_{KL}[q_\phi(z|x) || p(z)]$ 是KL散度(详情请看***),一种衡量两个分布相似性的数值,它可以确保潜在表示是高斯的,使得具有相似特征的数据点具有相似的高斯表示.

我们已经概述了潜在变量模型背后的一般思想,但我们仍然需要指定近似(变分)后验 $q_\phi(z|x)$ 和模型似然 $\log(p_\theta(x|z))$.近似后验的常见选择是分解高斯分布

$$q_\phi(z|x) = \prod_{j=1}^d N(z_j | \mu_j, \sigma_j^2)$$

其中 μ 和 σ 表示模型的均值和标准差.

总而言之,VAE是一种基于提取出 $p(z|x)$ 的潜在高斯表示 $q_\phi(z|x)$ 的均值与方差的编码器网络.VAE中的解码器,它使用高斯 $q_\phi(z|x)$ 的样本作为输入,根据分布 $p_\theta(x|z)$ 生成新样本.我们通过使用反向传播最大化ELBO来计算所有的参数,但是高斯分布是随机的,不可微的.因此需要在编码器的输出和解码器的输入之间建立一个确定性和可微分的映射.我们需要将随机变量 z 表示为另一个辅助随机变量 ϵ 的可微可逆变化 g (即 $z = g(\mu, \sigma, \epsilon)$).我们采用重参数化技巧,使 $g(\mu, \sigma, \epsilon) = \mu + \sigma \odot \epsilon$ 即

$$z = \mu + \sigma \odot \epsilon$$

其中 $\epsilon \sim N(0, 1)$ 和 \odot 是元素乘积.

自旋图变分自编码

我们想要构建一个图变分自编码器, 将VAE的思想应用于图结构数据. 我们希望我们的图变分自编码能生成新的不同尺寸的稳态下的Ising构型. 但是我们不能直接使用VAE, 因为传统卷积只能适用于同一尺寸下的Ising构型, 不同的Ising构型相当于不同的图结构, 节点数量不同导致不能直接使用卷积. 这时候我们就可以使用上文提到的图卷积, 利用图卷积可以有效描述不同的图结构, 将不同尺寸的构型放在同一个神经网络中训练, 并且得到我们想要的生成效果, 即生成出不同尺寸的稳态Ising构型.

编码器部分: 前两个GCN层生成一个高维特征矩阵,同时我们考虑图结构中的边信息,将边的特征参与到卷积聚合当中,参考之前图卷积的更新公式可以得到我们新的更新公式:

$$\mathbf{x}'_i = \mathbf{W}_1 \mathbf{x}_i + \mathbf{W}_2 \sum_{j \in \mathcal{N}(i)} e_{j,i} \cdot \mathbf{x}_j$$

其中 $e_{j,i}$ 表示从源节点 j 到目标节点 i 的边权重.

$$\hat{X} = GCN(GCN(X, A), A)$$

第三个卷积层我们采用两个不同的卷积网络来分别获取 $\mu, \log \sigma^2$

$$\begin{aligned} \mu &= GCN_\mu(\hat{X}, A) \\ \log \sigma^2 &= GCN_{\log \sigma^2}(\hat{X}, A) \end{aligned}$$

然后我们就可以从分布中进行采样 z :

$$z = \mu + \sigma \odot \epsilon, \epsilon \sim N(0, 1)$$

解码器部分:

我们的采样结果 z 是节点的高维表示,在解码器部分,由于Ising模型的节点之间的连接方式是固定的,所以我们在解码器部分可以忽略Ising模型的边属性,对潜在变量 z 中的每个节点使用全连接神经网络

络,使其维度降维至1.此时我们可以将重构问题认为是对每个节点进行分类,使用sigmoid激活函数,再将其按值进行分类.得到我们生成的重构模型

$$\overline{x}_i = MLP(z_i)$$

z_i 代表节点的高维表示,通过全连接神经网络使其变为我们需要的重构节点

损失函数:

$$L = MSE(X, \overline{X}) + KL(q(Z|X, A)||p(Z))$$

第一项MSE是均方误差,用于测量网络重构数据的程度,防止重构数据过度偏离原始数据;第二项KL散度,其中 $p(Z) = N(0, 1)$.它衡量我们的 $q(Z|X, A)$ 与 $p(Z)$ 的匹配程度

模型生成

在经过多次迭代以后,损失函数的值小于阈值后,我们就可以认为模型训练成功,这时候我们可以保存模型和训练出的 μ 与 σ 并使用解码器来生成我们想要的构型.

我们的模型可以生成多种尺寸的构型,只需要在训练时让你需要的尺寸放在一起进行训练,在训练完成后,我们再使用该尺寸下的构型将模型固定成我们需要的尺寸,就可以生成我们想要的尺寸下的构型

模拟结果与分析

为了验证上述网络的有效性,我们构建了一个Ising模型模拟器

总结