

骨关节炎风险预测模型构建

第二章我们获取了骨关节炎患者的基因型数据并对其进行了预处理与特征筛选。本章则主要对骨关节炎风险预测模型做以介绍。该模型首先根据基于变分期望最大化算法的图估计器构建了输入基因型数据的关联并通过图来表示该关联关系。之后构建了图神经网络对建立的图进行学习与管理。该模型还通过融合患者表型进一步增强模型性能。最后本模型还构建了基于预测值与输入数据的模型解释器。

图估计器

图神经网络中无结构数据的处理方法

图神经网络只适用于图类型的数据，对常见无结构数据（例如文本，图像）的处理较为困难。本文目前获得的基因型数据也属于无结构数据，原理上不适用于图神经网络。但是，考虑到图神经网络优秀的性能与可解释性，我们需要基于该无结构基因型数据构建图，继而输入图神经网络。目前根据无结构数据建图的方法主要有两种方法：一种是通过分析数据构建静态图，神经网络训练过程中不对该图结构加以更新[1][2]，例如在处理无结构的图像数据时，Monti[2]提出了一种通过超像素将图像数据转为图的方法，使得图神经网络能够处理图像数据；第二种是通过学习方法构建动态图结构，并在网络训练过程中不断更新图结构[3]。这类通过学习方法获取图结构特征的过程也被称为图结构学习（Graph Structure Learning, GSL）。目前的图结构学习过程主要分为三个步骤：首先通过已知数据通过k近邻法[4]或者阈值法[5]构建出一个初始图结构或观测图结构；然后使用诸如随机块模型（Stochastic Block Model, SBM）[6]对该观测图结构进行建模；最后通过算法估计出数据可能具有的图结构。在图估计的过程中，基于统计学的算法受到了广泛关注：例如Zhang[7]等人提出了一种基于贝叶斯思想的图估计器。其认为图结构根据依赖一定参数的分布所产生，如果通过蒙特卡洛方法对该参数进行估计，就可因此得到该分布的具体信息，继而对该图结构进行估计。Elinas[8]同样基于贝叶斯思想提出了一种使用变分推断法的图估计方法。这类贝叶斯方法思路简洁，原理清晰。因此，本文也将基于贝叶斯思想构建图结构估计器。

问题描述

贝叶斯思想认为：事件的观测值并不能反映事件的真实特性。贝叶斯方法因此主要解决由事件观测值向事件真实值的推断过程。但是在讨论推断过程前，我们首先需要对观测值与真实值加以定义。

观测值构建

对于本文研究对象基因型数据而言，未经图神经网络处理的原始数据所含信息量较少，较难依据初始数据构建观测值。但是有研究显示，经过谱图神经网络处理之后，具有较强关联的节点具有相似的值。[9]我们因此认为在图神经网络的输出值中，对任一节点，如果有一其他节点与其有着相似的值，则两节点之间可能存在关联。这种关联可以通过k近邻算法[4]计算从而得到k近邻网络，而该k近邻网络可以很好地描述节点的局部特征。在Wang[10]等人的研究中，也有通过根据图神经网络输出值构建k近邻网络的描述。我们因此根据每个样本经图神经网络处理后的输出值建立k近邻网络，再将所有网络相加求平均以获得单个观测值。同时为了防止信息丢失，我们将首次经图神经网络处理时输入的图初始化为全连接图，即认为每个节点都同其他节点存在关联。

真实值构建

本研究中所感兴趣的真实值为实际的未知图结构。但是对图结构的直接计算过于复杂，我们因此需要将图结构参数化，即使用少数几个参数来描述图结构。目前在图论领域常用的图参数化模型为随机块模型[6]，该模型认为图中的节点属于某几个簇，节点之间是否相连仅与两节点所在的簇相关且服从某一参数化分布。因此我们只需要得出支配该分布的参数就能对图结构加以复现，继而用少数参数描述庞大的网络。因此本文将该类参数作为推断的目标，而由其产生的图结构作为模型输出值。

推断过程

界定了观测值与真实值之后，我们便可对图估计器所解决的问题加以严谨描述。

图估计器首先根据图神经网络的输出构建观测值 O ，其中

$$O^{ij} = \frac{1}{N} \sum_N o_n^{ij}$$

O^{ij} 描述了节点 i, j 之间关联的观测值, N 为样本数, o_n^{ij} 为由第 n 个样本神经网络输出值构建的 k 近邻网络中节点 i, j 之间的关联。通过该观测值我们构建图 $G = (V, O)$, 其中 $V = \{v_1, \dots, v_n\}$ 描述图中节点, 即样本基因型位点。

为了通过SBM模型参数化图结构, 我们假定图中共含有 C 个簇, 每个节点 i 从属于簇 m 的概率用 $z_{im} \in \mathbf{Z}^{N \times C}, i \in \{1, \dots, n\}, m \in \{1, \dots, C\}$ 来描述, 且 z_i 相互独立并满足多项分布

$$z_i \stackrel{iid}{\sim} M(1, \alpha)$$

基于此, 我们使用 z_i 来描述节点之间关联。我们认为任意两节点之间的关联相互独立并满足二项分布

$$o_{ij}|z_{im}, z_{jn} = 1 \sim B(1, \pi_{z_i z_j})$$

可以看出, 只要解出参数 π 与 α , 就可以此构建出一个SBM模型, 继而对图结构加以推断; 因此本文给出以下目标:

基于观察到的节点关联信息 O 以及隐变量 Z , 估计参数 $\theta = \{\alpha, \pi\}$, 并根据该参数给出推断图结构

$$\hat{\theta} = \operatorname{argmax}_{\theta} l_{\theta}(O)$$

基于变分期望最大化的图估计

推导与整理

在节点关联与节点簇信息均已知的情况下, 我们首先计算似然函数 l_{θ} , 根据边缘概率公式, 有

$$l_{\theta}(O) = \sum_Z \log p_{\theta}(O, Z) = \sum_Z (\log p_{\pi}(O|Z) + \log p_{\alpha}(Z))$$

其中

$$\sum_Z (\log p_{\pi}(O|Z) + \log p_{\alpha}(Z)) = \sum_Z \left(\sum_{i,j} \sum_{m,n}^C \log p_{\pi}(o_{ij}|z_{im}z_{jn}) + \sum_i^N \sum_m^C \log p_{\alpha}(z_{im}) \right)$$

给定参数时, 有

$$p_{\pi}(o_{ij}|z_{im}z_{jn}; \pi) = \binom{1}{o_{ij}} [\pi_{mn}(1 - \pi_{mn})^{1-o_{ij}}]^{z_{im}z_{jn}}$$

$$p_{\alpha}(z_{im}) = \alpha_m^{z_{im}}$$

可得

$$\begin{aligned} l_{\theta}(O; \theta) &= \sum_Z \left(\sum_{i,j} \sum_{m,n}^C \log p_{\pi}(o_{ij}|z_{im}z_{jn}) + \sum_i^N \sum_m^C \log p_{\alpha}(z_{im}) \right) \\ &= \sum_Z \left(\sum_{i,j} \sum_{m,n}^C \log \left(\binom{1}{o_{ij}} [\pi_{mn}(1 - \pi_{mn})^{1-o_{ij}}]^{z_{im}z_{jn}} \right) + \sum_i^N \sum_m^C \log \alpha_m^{z_{im}} \right) \end{aligned}$$

期望最大化算法

考虑到似然函数中含隐变量 Z , 通常我们采用期望最大化算法[11]计算使得似然函数取极大的参数。我们对似然函数做如下处理

$$\begin{aligned} l(O; \theta) &= \log p(O; \theta) \\ &= \log \int_Z p(O, Z; \theta) dz \\ &= \log \int_Z q(Z) \frac{p(O, Z; \theta)}{q(Z)} dz \\ &= \log \mathbb{E} \left[\frac{p(O, Z; \theta)}{q(Z)} \right] \end{aligned}$$

其中 $q(Z)$ 为一辅助函数且满足 $\int q(Z) = 1$ 。根据Jensen不等式[14], 对任意一凸函数, 有

$$\mathbb{E}[f(x)] \leq f(\mathbb{E}[x])$$

当且仅当 $x = \mathbb{E}[x]$ 时等号成立。又因为 $\log x$ 为一凸函数，因此有

$$\begin{aligned} l(O; \theta) &= \log p(O; \theta) \\ &= \log \mathbb{E}\left[\frac{p(O, Z; \theta)}{q(Z)}\right] \\ &\geq \mathbb{E}\left[\log \frac{p(O, Z; \theta)}{q(Z)}\right] \end{aligned}$$

当且仅当

$$\frac{p(O, Z; \theta)}{q(Z)} = c$$

时取等（ c 为一常数）。对该式进行变化，有

$$q(Z) = c \cdot p(O, Z; \theta)$$

又因为 $\int q(Z) dz = 1$ ，有

$$\begin{aligned} \int_Z q(Z) dz &= \int_Z c \cdot p(O, Z; \theta) dz = 1 \\ c &= \frac{1}{\int_Z p(O, Z; \theta) dz} \end{aligned}$$

因此

$$q(Z) = \frac{p(O, Z; \theta)}{\int_Z p(O, Z; \theta) dz} = \frac{p(O, Z; \theta)}{p(O; \theta)} = p(Z|O; \theta)$$

代回原不等式，我们构造辅助函数 Q

$$l(O; \theta) \geq \mathbb{E}\left[\log \frac{p(O, Z; \theta)}{q(Z)}\right] = \int_Z p(Z|O; \theta) \log \frac{p(O, Z; \theta)}{p(Z|O; \theta)} dz = Q(\theta^{t+1}; \theta)$$

显然，在取等条件成立时， Q 为似然函数 $l(O; \theta)$ 的下界。期望最大化算法因此首先计算给定预期参数 θ 时下界 Q 的值，再通过优化参数 θ 提高下界 Q ，通过不断迭代计算使得似然函数极大的参数 $\hat{\theta}$ 。具体步骤描述如下

- E步骤：计算 $Q(\theta^{t+1}; \theta)$
- M步骤：优化参数 θ

但在下界 Q 的计算过程中我们设 $q(Z) = p(Z|O; \theta)$ ，但是在本问题中， $p(Z|O; \theta)$ 过于复杂，无法计算，因此也无法完成后续步骤。我们需要采取其他方法来估计参数 θ 。

变分期望最大化

变分推理

变分推理是贝叶斯统计领域遇到无法计算的分布时常见的处理方法[12]。它通过构建辅助函数来模拟无法计算的分布，并通过优化过程使得辅助函数逐渐逼近真实分布。[13]该方法与本文所处理问题较为契合，我们因此结合变分推理来处理期望最大化过程中无法计算的分布 $p(Z|O; \theta)$ 。

证据下限与KL散度

在进一步解释之前我们需要先对可能使用到的概念加以说明。在上一节中我们已经证明，根据Jensen不等式，含隐变量的似然函数具有一下界。我们将该下界定义为似然函数的证据下界(Evidence Lower Bound, ELBO)[15]。其具有如下形式

$$\begin{aligned} l(O; \theta) &= \log \int_Z q(Z) \frac{p(O, Z; \theta)}{q(Z)} dz \\ &\geq \int_Z q(Z) \log \frac{p(O, Z; \theta)}{q(Z)} dz \\ &= J(q(Z); \theta) \end{aligned}$$

我们同时定义似然函数与证据下界之间的差

$$\begin{aligned}
l(O; \theta) &= \int_Z q(Z) \log \frac{p(O, Z; \theta)}{q(Z)} dz \\
&= \int_Z q(Z) \log p(O; \theta) dz - \int_Z q(Z) \log \frac{p(O, Z; \theta)}{q(Z)} dz \\
&= \int_Z q(Z; \theta) \log \frac{p(O; \theta) q(Z; \theta)}{p(O, Z; \theta)} dz \\
&= - \int_Z q(Z; \theta) \log \frac{p(Z|O; \theta)}{q(Z; \theta)} dz \\
&= KL(q(Z) || p(Z|O))
\end{aligned}$$

为分布 $q(Z)$ 与分布 $p(Z|O)$ 之间的KL散度 (Kullback-Leibler divergence) [16]。可以看出, 该散度衡量了两分布间的差异, 可以用于后续辅助函数的构造。

辅助函数构建

基于变分推理思想, 我们构建辅助分布 $q_\psi(Z)$ 来模拟无法计算的 $p(Z|O)$ 。基于期望最大化思想, 我们构建如下似然函数

$$l(O; \theta) = J(q_\psi(Z); \theta) + KL(q_\psi(Z) || p(Z|O))$$

其中首项描述了在使用辅助分布替换 $p(Z|O)$ 时原始似然函数的下界, 第二项描述了辅助分布与原始分布之间的差值。值得注意的是, KL散度具有非负性, 即 $KL(\cdot || \cdot) \geq 0$ 。我们因此给出如下不等式

$$l(O; \theta) \geq J(q_\psi(Z); \theta)$$

借助期望最大化思想, 我们如果能够通过迭代计算辅助函数与参数使得该下界 $J(q_\psi(Z); \theta)$ 不断升高, 就能使得该似然函数最终取最大。展开该下界, 得

$$\begin{aligned}
J(q_\psi(Z); \theta) &= \int_Z q(Z) \log \frac{p(O, Z; \theta)}{q(Z)} dz \\
&= \int_Z q(Z) \log p(O, Z; \theta) dz - \int_Z q(Z) \log q(Z) dz \\
&= \int_Z q(Z) \left(\sum_{i,j}^n \sum_{m,n}^C \log \left(\frac{1}{O_{ij}} \right) [\pi_{mn}(1 - \pi_{mn})^{1-O_{ij}}]^{z_{im}z_{jn}} + \sum_i^N \sum_m^C \log \alpha_m^{z_{im}} \right) dz - \int_Z q(Z) \log q(Z) dz \\
&= \mathbb{E}_{q_\psi(Z)} \left[\sum_i^N \sum_m^C z_{im} \log \alpha_m \right] + \mathbb{E}_{q_\psi(Z)} \left[\sum_{i < j}^N \sum_{m,n}^C z_{im} z_{jn} \log \left(\left(\frac{1}{O_{ij}} \right) \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) \right] - \sum_i^N \sum_m^C \mathbb{E}_{q_\psi(Z)} [z_{im}] \log \mathbb{E}_{q_\psi(Z)} [z_{im}]
\end{aligned}$$

此时, 我们构造基于服从参数 τ 的多项分布的辅助函数, 使得

$$q_\psi(Z) = \prod_i^N q_\psi(Z_i) = \prod_i^N M(Z_i, \tau)$$

对于该变分参数 τ , 我们依据多项分布的性质给出如下限制条件 $\tau \in [0, 1]^C, \sum_m \tau_{im} = 1$ 。可以看出, 该参数同原始隐变量 Z 的意义相同, 都描述了节点属于簇的概率。因此我们给出

$$\begin{aligned}
\tau_{im} &= p(q_\psi(z_{im} = 1)) = \mathbb{E}_{q_\psi(Z)} [z_{im}] \\
\tau_{im} \tau_{jn} &= p(q_\psi(z_{im} = 1, z_{jn} = 1)) = \mathbb{E}_{q_\psi(Z)} [z_{im} z_{jn}]
\end{aligned}$$

代入原式, 有

$$J(O; \theta) = - \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_i \sum_m \tau_{im} \log \alpha_m + \sum_{i < j} \sum_{m,n} \tau_{im} \tau_{jn} \log \left(\left(\frac{1}{O_{ij}} \right) \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right)$$

此时我们可以发现, 虽然该下界的形式较为复杂, 但是在给定辅助函数与参数的情况下依然能够给出结果, 我们采用变分法的目的也就达到了。因此我们设计如下变分期望最大化算法

- VE步骤：固定参数 θ ，根据 J 更新变分参数 τ ，并计算更新参数后下界 J 的值
- M步骤：固定变分参数 τ ，根据 J 更新参数 θ ，直至收敛。

参数求解

在描述优化过程之后，我们对过程中参数的计算加以讨论

VE步

此步骤中主要求解变分参数 τ 。

$$\hat{\tau} = \arg \max_{\tau} J(O, \theta; \tau)$$

$$J(O, \theta; \tau) = - \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_i \sum_m \tau_{im} \log \alpha_m + \sum_{i < j}^N \sum_{m,n}^C \tau_{im} \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right)$$

考虑到待优化参数存在限制条件 $\sum_m \tau_{im} = 1$ ，我们采用拉格朗日乘数法加以计算

$$\begin{aligned} L(J, \lambda) &= J(O, \theta; \tau) + \lambda_i \left(\sum_m \tau_{im} - 1 \right) \\ &= - \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_i \sum_m \tau_{im} \log \alpha_m + \sum_{i < j}^N \sum_{m,n}^C \tau_{im} \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) + \lambda_i \left(\sum_m \tau_{im} - 1 \right) \end{aligned}$$

将该函数分别对 τ_{im} 与 λ 求偏导。得

$$\begin{aligned} \frac{\partial L}{\partial \tau_{im}} &= \sum_j^N \sum_n^C \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) + \log \alpha_m - \log \tau_{im} + \lambda_i - 1 = 0 \\ \frac{\partial L}{\partial \lambda_i} &= \tau_{im} - 1 = 0 \end{aligned}$$

有

$$\log \tau_{im} = \sum_j^N \sum_n^C \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) + \log \alpha_m - \lambda_i + 1$$

我们因此给出 τ_{im} 的解析解

$$\begin{aligned} \tau_{im} &= e^{-\lambda_i + 1} \alpha_m \left[\prod_j^N \prod_n^C \left[\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right]^{\tau_{jn}} \right] \\ \forall i &\in \{1 \dots N\}; \forall m \in \{1 \dots C\} \end{aligned}$$

可以看出，每个变分参数的计算都依赖于其他变分参数，虽然可以通过迭代至收敛的方法计算，但这也对参数初始值的选取提出了要求，具体细节将于后文讨论。

M步

此步骤中主要求解目标参数 θ 。

$$\hat{\theta} = \arg \max_{\theta} J(O, \tau; \theta)$$

首先我们计算 θ 中的 π 。对于

$$\begin{aligned} J(O, \tau, \alpha; \pi) &= - \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_i \sum_m \tau_{im} \log \alpha_m + \sum_{i < j}^N \sum_{m,n}^C \tau_{im} \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) \\ \frac{\partial J(O, \tau, \alpha; \pi)}{\partial \pi_{mn}} &= \sum_{i < j} \tau_{im} \tau_{jn} \left(\frac{O_{ij}}{\pi_{mn}} - \frac{1 - O_{ij}}{1 - \pi_{mn}} \right) \end{aligned}$$

可得

$$\pi_{mn}^{\hat{}} = \frac{\sum_{i < j} \tau_{im} \tau_{jn} O_{ij}}{\sum_{i < j} \tau_{im} \tau_{jn}}$$

对于 θ 中的 α ，由于其也存在限制条件 $\sum_m \alpha_m = 1$ ，我们依旧采用拉格朗日乘数法[17]计算。

$$\begin{aligned} L(J, \lambda) &= J(O, \pi, \tau; \alpha) + \lambda_i (\sum_m \alpha_m - 1) \\ &= - \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_i \sum_m \tau_{im} \log \alpha_m + \sum_{i < j}^N \sum_{m, n}^C \tau_{im} \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) + \lambda_i (\sum_m \alpha_m - 1) \end{aligned}$$

将该函数分别对 α_m 与 λ 求偏导。得

$$\begin{aligned} \frac{\partial L}{\partial \alpha_m} &= \frac{\sum_i \tau_{im}}{\alpha_m} + \lambda_i = 0 \\ \frac{\partial L}{\partial \lambda_i} &= \alpha_m - 1 = 0 \end{aligned}$$

对上式加以处理，得：

$$\begin{aligned} \frac{\sum_i \tau_{im}}{\alpha_m} + \lambda_i &= 0 \\ \sum_i \tau_{im} + \alpha_m \lambda_i &= 0 \\ \sum_m \sum_i \tau_{im} + \sum_m \alpha_m \lambda_i &= 0 \\ \lambda_i &= - \sum_m \sum_i \tau_{im} \end{aligned}$$

代入原式，得

$$\begin{aligned} \alpha_m^{\hat{}} &= \frac{\sum_i \tau_{im}}{\sum_m \sum_i \tau_{im}} \\ &= \frac{\sum_i \tau_{im}}{N} \end{aligned}$$

至此，我们已经完成了所有参数的估计。确定参数后只需将其代回原始模型即可得出估计的节点之间关联，继而完成图结构的构建。

起始与终止条件

之前在对变分参数 τ 的计算中我们发现，任一参数的计算都依赖于其他参数。但是如果以零为初始值，参数在计算时只会考虑项 $e^{-\lambda_i+1} \alpha_m$ ，无法对参数 π 形成有效估计。因此我们需要通过一定方法给予该变分参数初值，并通过迭代的方法计算 τ 直至收敛。目前的研究中常使用k-Means法[18]先对节点进行聚类，根据聚类结果对变分参数进行赋值。但是k-Means法需要预先确定图中节点簇数 C ，且对于一般问题而言该簇数未知，因此需要一种能够评估所选簇数对最终似然函数影响的方法。目前研究中，一种叫做贝叶斯信息指标（Bayesian Information Criterion, BIC）[19]的方法可以用来完成类似工作，其定义为。

$$BIC(C) = \log P(O; \theta) - \frac{V_C}{2} \log N$$

其中 $\log P(O; \theta)$ 描述对应模型下的似然函数， V_C 描述了选择对应簇数 C 时模型参数数量，但是该方法涉及到无法求解似然函数的计算，在本研究中无法实现。但是，基于类似的思想，Daudin等人提出另一种评价指标：整合分类似然（Integrated Classification Likelihood, ICL）[20]，该指标并不直接计算原始似然函数，而是通过计算变分时使用的变分似然函数给出评判依据。该指标定义如下

$$\begin{aligned} ICL(C) &= \sum_i \sum_m \tau_{im} \log \tau_{im} + \sum_{i < j}^N \sum_{m, n}^C \tau_{im} \tau_{jn} \log \left(\binom{1}{O_{ij}} \pi_{mn} (1 - \pi_{mn})^{1-O_{ij}} \right) \\ &\quad - \frac{1}{2} \left(\frac{C(C+1)}{2} \log \frac{N(N-1)}{2} - (C-1) \log N \right) \end{aligned}$$

基于以上信息，我们给出基于变分期望最大化的图估计算法全流程

- 输入：观测值，确定簇数范围
- 对于每簇数
 - 使用k-Means法对输入值进行聚类，给出初始变分参数 τ_0
 - VE步
 - 根据 $\hat{\tau} = \arg \max_{\tau} J(O, \theta; \tau)$ ，计算变分参数 τ 直至收敛
 - 根据该变分参数计算 J
 - M步
 - 根据 $\hat{\theta} = \arg \max_{\theta} J(O, \tau; \theta)$ 计算参数 θ
 - 迭代直至 θ 收敛
 - 收敛后计算最优似然下节 \hat{J} 与该簇数下的整合分类似然 $ICL(C)$
- 输出：选择使得整个分类似然最小的簇数 C ，选择该簇数计算出的参数 $\hat{\theta}$ 。基于该参数计算可能的图结构。

图神经网络

在完成了输入图结构的估计后，我们便可开始设计模型中图处理的核心——图神经网络。基于绪论中对图神经网络的讨论与综述，本文决定使用谱图神经网络中的切比雪夫层来处理生成的图数据。

基本定义

为了高效处理图数据，我们将图估计器所生成的节点关系通过矩阵表示，这类描述图节点之间关联的矩阵被称为邻接矩阵（Adjacency Matrix），用 A 来表示

$$A_{ij} = \{0, 1\}^{N \times N}$$

$A_{ij} = 1$ 意味着节点 i, j 之间存在相关，反之 A_{ij} 意味着两节点之间无相关关系。我们同时定义无向图中节点的度（Degree）为与该节点相关节点的数量并定义图的度矩阵 D 为

$$D_{ii} = \sum_{i \neq j} A_{ij}$$

可以看出，该度矩阵为一对角矩阵。同时为了后续矩阵的特征分解操作，我们定义图的拉普拉斯矩阵 L 为度矩阵与邻接矩阵之差，即

$$L = D - A$$

切比雪夫层

切比雪夫网络是一种谱图神经网络，其通过卷积方式来从数据中学习复杂信息[21]。但是同常见的卷积神经网络不同，谱图神经网络通过傅里叶变换的方式来实现图这种非欧数据上的卷积。根据卷积定理[22]，两个函数的卷积为他们傅里叶变换后的结果的点积的傅里叶逆变换，即

$$x * y = F^{-1}\{F\{x\} \cdot F\{y\}\}$$

因此如果我们能够在图上定义一傅里叶变换，我们就能以此定义图数据上的卷积。通常来说，函数的傅里叶变换是函数在拉普拉斯算子特征函数这一组标准正交基上的投影。类似地，我们根据图的拉普拉斯矩阵定义该变换。

$$Lu = \lambda u$$

其中 L 为图的拉普拉斯量， u 为该正交基， λ 为特征值。我们同时定义矩阵 $U = [u_1, \dots, u_N]$ ，使得

$$L = U^T \Lambda U$$

因此我们定义图 ϕ 上的傅里叶变换为

$$\begin{aligned} \hat{\phi} &= U^T \phi = \text{diag}(\hat{\phi}(\lambda_l)) \\ &= \text{diag}(\sum_i^N \phi_i u_{li}) \end{aligned}$$

因此根据卷积定理，对于滤波器 g 与图信号 x 而言，在图 G 上定义的卷积为

$$\begin{aligned}(g * x)_G &= U(U^T g \cdot U^T x) \\ &= U g_\theta(\Lambda) U^T x \\ &= g_\theta(L) x\end{aligned}$$

其中 g_θ 为卷积核在图中的傅里叶变换， θ 为该卷积核的参数。因此我们可以定义谱图神经网络中的一层为

$$y = \sigma(g_\theta(L)x)$$

但是，此时 g_θ 较难确定。Defferend[21]等人依靠切比雪夫多项式对 g_θ 进行了近似，其认为

$$g_\theta(\Lambda) = \sum_{k=0}^K \theta_k T_k(\tilde{\Lambda}), \quad \tilde{\Lambda} = 2\Lambda_n / \lambda_{max} - I_n$$

其中 θ 代表需要在训练中学习的参数， T_k 代表 k 阶切比雪夫行列式， $\tilde{\Lambda}$ 代表特征矩阵。其中切比雪夫行列式计算方式如下。

$$\begin{aligned}T^{(0)} &= X \\ T^{(1)} &= \tilde{L}X \\ T^{(k \geq 2)} &= 2 \cdot \tilde{L}T^{(k-1)} - T^{(k-2)},\end{aligned}$$

因此切比雪夫网络中的一层表示为

$$Y = \sigma \left(\sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x \right)$$

结构设计

基于切比雪夫层的特点，我们构建如下图神经网络。

- 切比雪夫层
- 切比雪夫层
- 读出层
- 全连接层

其中切比雪夫层设置阶数 $k = 1$ ，原理上该型卷积核同CNN中最经典的 3×3 卷积核作用结果相仿。每层切比雪夫层会使节点融合与其相邻节点的信息，增强模型对局部信息的处理能力。读出层将进行卷积操作的图再次转化为无结构数据，再将该数据输入全连接层中用以后续的风险预测。

表型融合

考虑到骨关节炎作为一种复杂多基因疾病，单纯以基因型为输入可能导致风险预测效果不佳。因此有必要同时融合表型数据。根据Boer[23]等人的报道，我们选择同骨关节炎密切相关的表型做为协变量并从UKB数据库中获取相关表型数据。具体表型如下表。

UKBID	表型
32883-31	性别
32883-48	髀围
32883-49	腕围
32883-2463	是否发生过骨折
32883-21001	BMI
32883-21002	体重
32883-21003	年龄

同时我们选择多模态数据融合中的中间融合法将基因型数据与表型数据相融合。我们将已构建的图神经网络的中间输出层与所获取的表型数据相融合，并输入新的分类器进行训练，最终给出融合表型后的预测结果。

图解释器

图是一种包含了节点信息与节点之间关系的数据结构，但是，本文目前为止建立的模型只能直接给出对患病风险的预测，与传统机器学习方法无异。如果能够充分利用图中的关联信息并对产生该预测结果的原因加以解释，不仅能够增强预测结果可信性，还能根据其解释内容发掘数据中潜藏的其他信息。目前对于图神经网络解释器的研究已有颇多进展，本文基于其中备受关注的GNNExplainer [24]来构建模型解释器。

GNNExplainer通过输入的已训练图神经网络模型及其预测结果从输入数据中选择对预测结果影响最大的子图。其数学描述如下

$$\max_{G_s} MI(Y, G_s) = H(Y) - H(Y|G = G_s)$$

其中

- MI 表示两变量之间的互信息（Mutual Information），其代表两变量之间相互依赖的程度。
- Y 表示模型给出的预测结果
- G_s 表示同预测结果相关的子图
- H 表示变量的信息熵

该文献描述：对于给定的GNN模型， $H(Y)$ 恒定，此时只需要最小化 $H(Y|G = G_s)$ 就可达到解释目的。该文献同时描述了一种变分近似方法来实现该优化过程，最终生成了与预测结果相关的最小子图。

本章小结

本章描述了骨关节炎风险预测模型的构建过程与基本模块。本文首先针对现存基因型数据结构缺失的问题基于变分期望最大化算法构建了一个图结构估计器。该估计器能根据神经网络的输出动态更新与预测图结构，同时能够对图中的节点聚类分析。其次，本文根据谱图神经网络中的切比雪夫层构建了包裹卷积层、读出层、全连接层在内的图神经网络用于图数据的处理以及患病风险的预测。考虑到单纯通过基因型数据预测患病风险结果可能较差，本文还基于文献描述的同骨关节炎密切相关的表型构建了表型融合风险预测模型。最后，为了尽最大可能挖掘出图数据中的潜藏信息，本文还根据已有的图神经网络解释器构建了预测结果解释器，在给出风险预测结果的同时获取同预测结果相关的子图。以上模块共同构成了本文预期的骨关节炎风险预测模型。

- [1]:Jiafeng Hu, Reynold Cheng, Zhipeng Huang, Yixang Fang, and Siquang Luo. 2017. On Embedding Uncertain Graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. Association for Computing Machinery, New York, NY, USA, 157–166. <https://doi.org/10.1145/3132847.3132885>
- [4]: [Preparata and Shamos, 1985] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer, 1985.
- [5]:[Bentley et al., 1977] J. L. Bentley, D. F. Stanat, and E. H. Williams Jr. The Complexity of Finding Fixed-Radius Near Neighbors. *Inf. Process. Lett.*, 1977.
- [6]: Holland, Paul W; Laskey, Kathryn Blackmond; Leinhardt, Samuel (1983). "[Stochastic blockmodels: First steps](#)". *Social Networks*. **5** (2): 109–137. [doi:10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7). [ISSN 0378-8733](#). Retrieved 2021-06-16.
- [7]:Y. Zhang, S. Pal, M. Coates, and D. Üstebay. Bayesian Graph Convolutional Neural Networks for Semi-Supervised Classification. In *AAAI*, 2019.
- [8]:[Elinas et al., 2020] P. Elinas, E. V. Bonilla, and L. C. Tiao. Variational Inference for Graph Convolutional Networks in the Absence of Graph Data and Adversarial Settings. In *NeurIPS*, 2020.
- [10]:Ruijia Wang, Shuai Mou, Xiao Wang, Wanpeng Xiao, Qi Ju, Chuan Shi, and Xing Xie. 2021. Graph Structure Estimation Neural Networks. In *Proceedings of the Web Conference 2021(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 342–353. <https://doi.org/10.1145/3442381.3449952>
- [11]:Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39, 1–38.
- [12]:Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37, 183–233.
- [13]:Jaakkola, T. S., and Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10, 25–37.
- [14]: Jensen, J. L. W. V. (1906). "[Sur les fonctions convexes et les inégalités entre les valeurs moyennes](#)". *Acta Mathematica*. **30** (1): 175–193. [doi:10.1007/BF02418571](https://doi.org/10.1007/BF02418571)
- [15]:D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, The Variational Approximation for Bayesian Inference, *IEEE Signal Processing Magazine*, Nov 2008
- [16]:R.M. Neal and G.E. Hinton, A view of the EM algorithm that justifies incremental, sparse and other variants, *Learning in Graphical Models*, 1998
- [17]:Abir El Haj. Stochastics blockmodels, classifications and applications. Modeling and Simulation. Université de Poitiers; Université Libanaise, 2019. English. ffnnt : 2019POIT2300ff. fftel-02926379f
- [18]:Abir El Haj, Yousri Slaoui, Pierre-Yves Louis, Zaher Khraibani. Estimation in a Binomial Stochastic Blockmodel for a Weighted Graph by a Variational Expectation Maximization Algorithm. *Communications in Statistics - Simulation and Computation*, Taylor & Francis, 2020, [ff10.1080/03610918.2020.1743858](https://doi.org/10.1080/03610918.2020.1743858)ff. fahal-02271491v1f
- [19]: Schwarz, Gideon E. (1978), "Estimating the dimension of a model", *Annals of Statistics*, **6** (2): 461–464, [doi:10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136), [MR 0468014](#).
- [20]:Daudin, J. J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and computing*, 2, 173–183.

[22]: Proakis, John G.; Manolakis, Dimitri G. (1996), [*Digital Signal Processing: Principles, Algorithms and Applications*](#) (3 ed.), New Jersey: Prentice-Hall International, p. 297, [Bibcode:1996dspp.book.....P](#), [ISBN 9780133942897](#), sAcfAQAAIAAJ

[23]:Boer, C. G., Hatzikotoulas, K., Southam, L., Stefánsdóttir, L., Zhang, Y., Coutinho de Almeida, R., Wu, T. T., Zheng, J., Hartley, A., Teder-Laving, M., Skogholt, A. H., Terao, C., Zengini, E., Alexiadis, G., Barysenka, A., Bjornsdottir, G., Gabrielsen, M. E., Gilly, A., Ingvarsson, T., . . . Zeggini, E. (2021). Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*, 184(18), 4784-4818.e4717. <https://doi.org/10.1016/j.cell.2021.07.038>

[24]:Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23(2), bbab569. <https://doi.org/10.1093/bib/bbab569>