年前就希望系统的学习下概率图模型，加上最近新模型的设计缺乏灵感，为了找一些想法，翻看了概率图模型的基本理论，还有概率图模型在 GNN 领域中应用的几篇顶会论文。

基本理论（这部分参考了 “周志华-机器学习” 及 “概率图模型” 两书）

概率图模型是指用图的方式来表示变量之间的概率分布，所以图在概率图模型中充当了一个工具。概率图模型分为两类：1）使用有向无环图表示变量间的依赖关系，成为贝叶斯网（变量间存在因果关系进而有向）；2）使用无向图表示变量间的依赖关系，成为马尔可夫网（变量间存在相关但不知道因果关系）。

这说明了概率图模型和 GNN 很不一样的地方，概率图模型是以图为工具刻画变量间的依赖关系。GNN 是关注如何处理图数据，学习节点或图的表达，或者解决图数据上的任务等。图数据中的连边可能是物理世界产生的，如社交网络、引用网络中的连边，因此连边不一定表示变量间的相关关系。现有工作在 GNN 中引入概率图模型，其实是把概率图模型作为实现 GNN 中某个约束的工具。

隐马尔可夫模型

隐马尔可夫模型是一个链式有向图模型，因此属于贝叶斯网。隐马尔可夫中变量分为两类，状态变量（不可观测）和观测变量（可观测）。隐马尔可夫有两个性质：1）当前观测只依赖于当前状态；2）当前状态只依赖于上一个状态。

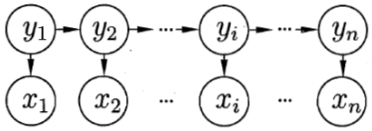


图 隐马尔可夫模型的图结构

因此所有变量的联合概率分布为：

$$P(x_1, y_1, \ldots, x_n, y_n) = P(y_1)P(x_1 \mid y_1)\prod_{i=2}^{n} P(y_i \mid y_{i-1})P(x_i \mid y_i)$$

确定一个隐马尔可夫模型需要三组参数：1）状态转移概率；2）输出观测概率；3）初始状态概率。也有三个比较关心的问题：1）在给定当前观测序列下，下一个概率最大的观测状态是什么（文本生成）；2）当前观测序列对应的最可能的状态序列是什么（语音识别）；3）在给定观测数据下，最好的模型参数是什么。

发散：在图数据的节点分类问题下，是不是可以把节点的 feature 作为观测变量，把节点的 label 作为状态变量，我们要解决的问题就是在给定观测下最可

能的状态是什么。我们可以假设这个问题满足隐马尔可夫的第一个性质，即当前观测仅和当前状态相关，但是图数据不是一个链式模型，因此当前状态不仅和上一个状态相关，应该考虑的是无向图（非有向链）上的相关性。

生成式模型，考虑联合分布

马尔可夫随机场

隐马尔可夫的约束比较强，要求当前状态只依赖于上一个状态，但实际上，变量之间可能仅知道相关性，不知道因果关系，且变量间可能存在复杂的相关性，因此马尔可夫随机场将链式模型扩展到图上。

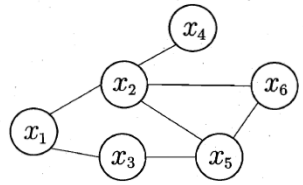


图 14.2　一个简单的马尔可夫随机场

马尔可夫随机场是无向图模型，有一组势函数（因子，非负函数），用来定义概率分布。势函数是定义变量之间相关性的非负函数，势函数并不是一个概率函数，因此为了使得概率取值在 0-1 之间，势函数到概率分布函数的转化需要归一化。在无向图模型，联合概率分布函数的分解是基于最大团进行的

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{Q \in \mathcal{C}} \psi_Q(\mathbf{x}_Q) \ ,$$

其中 Q 是图上的最大团，ψ_Q是 Q 上的势函数，指示了 Q 中的变量相关性。这个部分我还没完全理解，因为不同 Q 中的节点是有交集的，所以不同 Q 是有相关性的，但是势函数中仅定义了 Q 内的变量相关性，没有考虑 Q 间的变量相关性。对于两个不同的 Q，取出他们的交集外，剩余的部分相互独立。

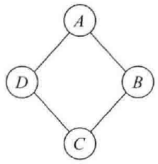
习题 2.5

令 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ 是三个不相交的变量子集，且 $\mathcal{X}=\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z}$。证明：$P \vDash (\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z})$ 当且仅当 P 有如下形式：

$$P(\mathcal{X}) = \phi_1(\boldsymbol{X}, \boldsymbol{Z}) \phi_2(\boldsymbol{Y}, \boldsymbol{Z})$$

或者是这样理解的，在一个新的问题下，P 是不知道的，但是是与节点相关性有关的。用势函数定义变量之间的相关性，针对希望出现的状况给大的值，然后做归一化使得其结果满足概率分布的约束。概率图模型中有个关于势函数的例子，为了有助于理解，我放在下面：

例 3.8 考虑有四个学生的情形，这些学生要求两两一组来完成课程作业。由于各种原因，只有如下几对组合适合：Alice 和 Bob；Bob 和 Charles；Charles 和 Debbie；以及 Debbie 和 Alice。(Alice 和 Charles 相互之间合不来，而 Bob 和 Debbie 刚刚结束了一段糟糕的恋情。)这些学习组合如图 3.10(a)所示。



(a)

这个题目里，最大团为{A,B},{A,D},{B,C},{C,D},每个最大团上的势函数就是定义团内变量的相关性，势函数的定义可以如下：

$\phi_1(A, B)$			$\phi_2(B, C)$			$\phi_3(C, D)$			$\phi_4(D, A)$		
a^0	b^0	30	b^0	c^0	100	c^0	d^0	1	d^0	a^0	100
a^0	b^1	5	b^0	c^1	1	c^0	d^1	100	d^0	a^1	1
a^1	b^0	1	b^1	c^0	1	c^1	d^0	100	d^1	a^0	1
a^1	b^1	10	b^1	c^1	100	c^1	d^1	1	d^1	a^1	100
(a)			(b)			(c)			(d)		

图 4.1 误解示例的因子

上述势函数在直接相关的变量之间定义了局部交互影响，为了定义全局概率分布，马尔可夫随机场对这些因子以相乘的方式组合。为了满足概率分布的形式，又做了归一化，结果如下：

赋值				非归一化	归一化
a^0	b^0	c^0	d^0	300, 000	0.04
a^0	b^0	c^0	d^1	300, 000	0.04
a^0	b^0	c^1	d^0	300, 000	0.04
a^0	b^0	c^1	d^1	30	$4.1 \cdot 10^{-6}$
a^0	b^1	c^0	d^0	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^0	d^1	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^1	d^0	5, 000, 000	0.69
a^0	b^1	c^1	d^1	500	$6.9 \cdot 10^{-5}$
a^1	b^0	c^0	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^0	d^1	1, 000, 000	0.14
a^1	b^0	c^1	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^1	d^1	100	$1.4 \cdot 10^{-5}$
a^1	b^1	c^0	d^0	10	$1.4 \cdot 10^{-6}$
a^1	b^1	c^0	d^1	100, 000	0.014
a^1	b^1	c^1	d^0	100, 000	0.014
a^1	b^1	c^1	d^1	100, 000	0.014

若 A 变量到 B 变量必须经过 C 变量集，则 C 称为 AB 的分离集。马尔可夫随机场借助分离集满足三个条件，全局马尔可夫性，局部马尔可夫性，成对马尔可夫性。

马尔可夫随机场也是建模联合概率分布，属于生成式模型。

条件随机场

前面还有最大熵马尔可夫模型，区别在于最大熵马尔科夫用局部归一化，所以存在标注偏置问题。例如 :当某个节点的转移状态较多时，在这个节点归一化，使得这个节点即便到倾向到达的状态，概率也比较低。但如果一个节点的转移状

态较少，在做了节点归一化后，这个节点到其他状态的转移概率都比较大。

生成式模型都是在建模联合概率分布，判别式模型则是在给定条件下，建模条件概率。这里没有 get 到判别式模型和生成式模型是否有优劣，但是判别式模型是不是需要更多数据才能训练好？因为判别式模型是不是需要考虑在每个给定特征下的推断概率？

条件随机场也是无向图模型，属于判别式模型，以分类问题为例，条件随机场直接寻找分类边界。

> 令 $G = \langle V, E \rangle$ 表示结点与标记变量 $\mathbf{y}$ 中元素一一对应的无向图，y_v 表示与结点 v 对应的标记变量，$n(v)$ 表示结点 v 的邻接结点，若图 G 的每个变量 y_v 都满足马尔可夫性，即
>
> $$P(y_v \mid \mathbf{x}, \mathbf{y}_{V\setminus\{v\}}) = P(y_v \mid \mathbf{x}, \mathbf{y}_{n(v)}) \quad , \tag{14.10}$$
>
> 则 $(\mathbf{y}, \mathbf{x})$ 构成一个条件随机场.

和马尔可夫随机场相似，条件随机场也使用团上的势函数定义概率，只是条件随机场建模的是条件概率，而马尔可夫随机场建模的是联合概率。

概率图模型在 GNN 中的应用

我现在着重看了两篇，以我自己的理解，这两篇的模式比较像，都是说现在的 GNN 缺乏什么样的性质保证，然后通过自己定义势函数满足这个性质，最大化马尔可夫随机场的联合概率或者条件随机场的条件概率，使得这种势函数被满足。

Graph Convolutional Networks Meet Markov Random Fields: Semi-Supervised Community Detection in Attribute Networks（AAAI2019）

这篇写 GNN 在应用社区发现时具有很好的特征提取能力，但是 GNN 的设计不是为了社区发现，缺少社区发现的性质约束，比如势函数定义为倾向于语义相似和结构相近的节点同属一个社区，MRF 中概率分布定义成势函数的乘积，然后最大化概率分布。

Conditional Random Field Enhanced Graph Convolutional Neural Networks（KDD2019）

GNN 没有约束相连的节点隐层表达相似，这篇论文先通过结构和属性定义节点相关性，然后势函数要求相关性高的节点对，中间层表达相似。