This is a note for the paper:Graphical-BasedLearningEnviron(2004) && graph neural network model(2009)

问题背景

一个输入数据是一张图，图上有若干节点。可以是有向图，可以是有环图。节点带有特征向量、边上也带有特征向量。节点的特征向量表征节点的属性，边的特征向量表征节点之间的关系。问题可以是对节点分类，也可以是对图分类（分类可以被替换成回归）。不同输入数据的图拓扑可以不同，节点数量也可以不同。

问题的一个实例：子图匹配。寻找在图中特征子图，如果节点属于子图则输出为1，否则为-1。子图是固定的，但是输入的图可以各式各样。子图节点有特征，特征被加了噪声（因此不能精确匹配，但加个阈值也可以模糊匹配）。

一个实例：化学分子分类，能引起诱变或不能。输入是分子结构，包括原子属性和原子之间的键。不同分子结构和原子不同。

一个实例：网页打分。输入是节点标签特征，节点属性值，节点连接关系。输出是一个与三者有关的函数，相当于是学习这个函数。图是固定的，一个大图中有5000个节点。50个用作训练，50个用作验证，剩下大部分用于测试。

网络结构

作者希望在图上定义某种迭代运算，使图的状态最终收敛。为了保证收敛，作者要求这种迭代为压缩映射。若进一步约束输出为收敛状态的线性可微函数（从收敛状态到输出的映射是简单的，可以随便设计），作者证明了输出为输入及参数的连续可微函数（这并不是显然的，因为一般来说参数变一点，收敛点可能会发生跳变），因此可以采用梯度下降来拟合参数。作者的证明有一点问题，作者认为矩阵范数大于0则可逆，这是不对的。可以利用如下引理更改：

$$||A|| < 1, ||I|| = 1, \Rightarrow (I - A)\text{可逆，且}(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$$

从这个定理的证明中，我们可以看出，压缩映射定义了一个参数、拓扑结构和标签到输出状态的隐函数（注意：收敛的状态和输入是没关系的）。因此，如果输入数据是一张固定的图，节点及标签也固定，则输出是固定的，模型没有随机性，比如网页打分的例子。如果拓扑结构变化，则压缩映射可以动态变化。

转态转移

参考random walk, 每个节点的状态由父亲节点转态的线性（或非线性）变换的和（或者均值）得到。这个变换被记为$H_w(u)$ ，u为父亲节点。当前节点n的状态$x_{t+1}(n) = \sum_{u \in Parent[n]} H_w(u_t)$

迭代等价于在时间上展开的RNN，连接关系为图上节点的连接关系。

简单地，H可以由一个多层神经网络实现，为了使整体映射$f_w(x, l)$为压缩映射$||\frac{\partial f}{\partial x}||_1 < 1$，可以将约束转化为神经网络的损失函数$||\frac{\partial f}{\partial x}||_1$ 。若觉得一范数不太好求，也可以转化为所有元素的绝对值的和。

如果H是一个线性映射，那么线性映射的系数要由一个神经网络确定，要求满足$||\frac{\partial f}{\partial x}||_1 < 1$。这也很简单，只需保证每个元素绝对值小于1（激活函数限制，比如tanh），最后再除以矩阵维数即可。Bias可以由另一个神经网络确定。注意，Bias对收敛状态的影响是很大的。

输出函数

收敛状态到输出的函数，随意设计，可以是含一个隐层的全连接网络。

反向传播

作者推导了一种快速反向传播的方法。详细推导见论文，一步一步看即可。