

输入数据的格式

符号表

符号	含义
V	整个图的点集
V_{train}	训练集
V_{test}	测试集
V_s	有监督的点集
V_u	无监督的点集
N	整个图的节点数
N_{train}	训练集节点数
N_{test}	测试集节点数
N_s	有监督的节点数
N_u	无监督的节点数
F	每个节点的特征数
C	分类的类数（此问题中是2）

点集的划分

所有节点被划分为训练集和测试集，即 $V = V_{train} \cup V_{test}$ 。训练集用于训练模型，测试集用于评估模型。

其中训练集中节点被划分为有监督和无监督，即 $V_{train} = V_s \cup V_u$ 。有监督的节点数应当远少于训练集的总节点数，即 $N_s \ll N_{train}$ 。所有节点都有标签，无监督的节点和测试集的标签用于评估模型。

输入的描述

输入包含x, tx, allx, y, ty, ally, graph, test.index，它们可以被分为：

1. 特征矩阵
2. 标签矩阵
3. 邻接矩阵
4. 索引集合

特征矩阵

特征矩阵的每一行表示一个节点，每一列表示一个特征。
所有特征矩阵的类型均为 `scipy.sparse.csr.csr_matrix`。

1. x是一个 $N_s \times F$ 的矩阵，表示有监督节点的特征矩阵
2. tx是一个 $N_{test} \times F$ 的矩阵，表示测试集的特征矩阵
3. allx是一个 $N_{train} \times F$ 的矩阵，表示训练集的特征矩阵

标签矩阵

标签矩阵的每一行表示一个节点，每一列表示一个分类。

所有标签矩阵的类型均为 `numpy.ndarray`。

1. y 是一个 $N_s \times C$ 的矩阵，表示有监督节点的标签矩阵
2. ty 是一个 $N_{test} \times C$ 的矩阵，表示测试集的标签矩阵
3. $ally$ 是一个 $N_{train} \times C$ 的矩阵，表示训练集的标签矩阵

邻接矩阵

邻接矩阵的 (i, j) 号元素表示节点 i 与节点 j 之间的边权。

邻接矩阵的类型为字典。

`graph` 是一个 $N \times N$ 的矩阵，表示整个图的邻接矩阵。

索引集合

`test.index` 是一个包含 N_{test} 个元素的集合，表示测试集中所有节点的索引值。

存放方式

所有的输入文件按照上述命名，存放在 `gcn/data` 目录下。

其中特征矩阵、标签矩阵、邻接矩阵用 `pickle` 存放，`test.index` 直接将数据用换行符分隔存放。