

GKD：用于图无关推理的半监督图知识提取

摘要

摘要中提到了医学数据的多模态化增加了对疾病预测领域的综合洞察机会，使用图卷积网络 (GCNs) 的最新研究提供了集成异构模态的半监督方法来研究患者和疾病预测之间的关联。然而，当用于图构建的元数据在推理时不可用时，传统方法的性能表现较差。因此，作者提出了一种基于知识蒸馏的半监督方法GKD，通过在训练阶段使用标签传播算法以及深度神经网络的教师组件来嵌入图和非图模态的所有信息到软伪标签中，然后使用软伪标签来训练一个深度学生网络进行对无法得到图模态的未知测试数据的疾病预测。实验使用了两个公共数据集对自闭症谱系障碍和阿尔茨海默病进行诊断，以及对合成多模态数据集的彻底分析。根据实验结果，与以前的基于图的深度学习方法相比，GKD在准确性、AUC和 Macro F1等指标方面表现更佳。

介绍

文章介绍部分提到，对于医学图像和非医学图像数据的分析，需要一种集成框架。而随着传统深度神经网络的成功，越来越多的研究致力于将其应用于多模态数据集的分析。通过将患者作为节点、将它们之间的关联作为边来构建图模型，可以有效地分析人群之间的交互。基于图的卷积神经网络 (GCNs) 提供了一种深度网络架构，可以将节点特征和图模态数据进行集成，用于人群级别的分析。同时，GCN被应用于医疗应用中，包括疾病分类、脑部分析、乳腺X射线分析等。然而，在实践中，由于图模态数据不能完全获取，这会对模型的泛化能力产生影响。为了解决这个问题，提出了一种基于知识蒸馏技术的新的学生-教师方法，以在训练时利用所有模态，在推理时不受图模态的影响。该方法使用标签传播算法将所有与图相关的数据进行缩放，生成由所谓的教师模型产生的伪标签，然后用这些伪标签来训练一个学生网络，该网络在整个网络中不需要过滤特征即可将高方差和嘈杂的原始输入空间映射到输出。因此，该学生网络可以在未来的未知数据上进行预测，而不需要图像模态的存在。

知识蒸馏框架

复杂模型或由多个单独训练的模型集成的方法可以从数据中提取复杂的结构，但是在部署时效率不高。训练的模型的知识保存在其参数中，而在更改目标模型的结构、形式和大小时传递学到的权重并不直观。然而，训练的模型知识可以被视为从其输入到输出的映射。在分类任务中，训练的模型输出的类概率的相对大小是一种丰富的信息源，表示类之间的相似性。由训练好的教师网络获得的这种相似性度量可以被用作训练更小、更高效的学生网络。Caruana等人[4]和Hinton等人[13]的研究表明，学生模型的推广方式与教师模型相同，并且比从头开始使用硬目标进行训练的模型更优。受到这一框架的启发，论文中使用了所有的可用模态仅在教师的学习过程中，然后将获得的知识传递到一个没有与图形可用性相关的学生上。

方法

1. 问题定义和符号

- 图表示: $G(V, E, A)$ A 为节点的邻接矩阵 D 为 A 对角矩阵
- X 为输入 $X \in R^{N \times F}$ 为节点的特征维度
- Y 为节点的输出, $Y \in R^{N \times C}$
- 半监督学习——预测无标签样本的类别概率
 - 有标签 V_L : 真实标签 $Y_L = V_L$
 - 无标签 V_U : 预测类别概率

2. 教师模型

- 预测无标签样本的类别概率，并将这些样本与标记样本一起作为知识来训练学生模型

- 首先将标记样本训练DNN (D_θ^T) 并用训练后的DNN预测无标记样本得到最可能的软伪标签 \hat{Y}_U
- 将软伪标签与真实标签结合来初始化图
- 使用LPA (标签传播) 算法将图信息添加到节点标签中

$$Y^{T^{(k)}} = (1 - \alpha)D^{-1}AY^{T^{(k-1)}} + \underbrace{\alpha}_{\text{Remembrance term}} Y^{T^0} ; Y^{T^0} = Y_L \cup \hat{Y}_U,$$

$$Y_L^{T^{(k)}} = Y_L,$$

3. 学生模型

- 模拟教师网络的输入输出映射, 将原始图输入DNN中得到节点的预测概率 Y^S
- 通过最小化 Y^S 和 Y^T 的距离来更新学生模型 D_θ^S
- **距离**: KL散度、交叉熵损失

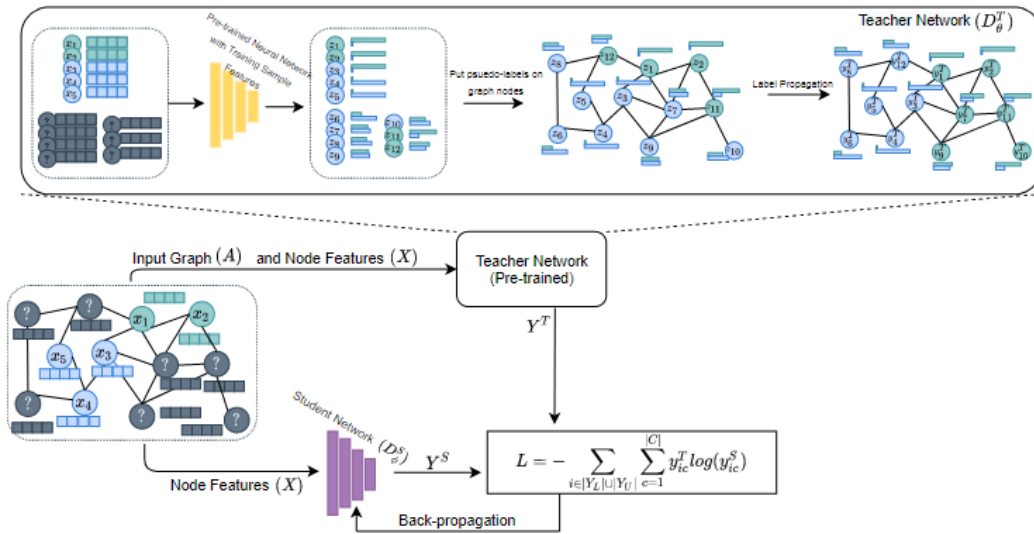


Fig. 1: Overview of the proposed method. The training set comprises of labeled and unlabeled nodes. Unlabeled nodes are in dark grey and other colors depicts the node labels. The upper part: teacher network. The lower part: student network. The student is trained by cross-entropy loss between teacher and student outputs to inject the teachers knowledge into the student's parameters.

实验和结果

实验设置

- 三个多模态医学数据集
 - 自闭症谱系障碍 (ABIDE)——自闭症脑成像数据交换 (ABIDE) [8, 6]数据库提供了1112名患者的神经成像 (功能性MRI) 和表型数据, 这些数据带有二进制标签, 表明存在诊断的自闭症谱系障碍 (ASD)
 - 阿尔茨海默病神经成像计划 (ADNI)——阿尔茨海默病纵向进化预测 (TAPOLE) [22]是阿尔茨海默病神经成像倡议 (ADNI) 数据的一个子集, 该数据集具有二进制标签, 指示患者的阿尔茨海默病状态是否会在未来六个月内得到改善。该数据集包含151例阿尔茨海默病诊断或进展的病例和1878例保持其状态的病例
 - 多模态合成数据集——生成2000个样本, 平均分布在两个类别中, 具有128维节点特征和4维图模态特征。节点特征生成算法来自[12], 该算法基于创建超立方体上正态分布在顶点周围的类点[25]。图模态的特征是从4维标准正态分布中提取的, 每个类的平均值分别为 $\mu c1$ 和 $\mu c2$ 。通过P缺失率随机去除图的特征值, 并遵循TADPOLE图构建的步骤。

- 网络架构: EV-GCN[15]和Parisot等人[23]引入的方法, **全连接神经网络**是教师和学生实验中选择的结构
- 参数设置: 隐层个数 (1、2、3) ; 学习率: {5e-3, 1e-2}; Dropout: {0.1, 0.3, 0.5}
- 优化算法: Adam算法

1. 自闭症谱系障碍 (ABIDE)

为了与其他方法进行公平的比较, 选择了同样的871名患者, 包括468名正常患者和403名ASD患者。还遵循Parisot等人[23]的相同预处理步骤, 从fMRI中选择2000个特征作为节点特征。分别使用55%、10%和35%用于训练、验证和测试, 并且标签可用于40%的训练样本。对于图的构建, 首先, 丢弃了仅适用于ASD患者的表型特征, 以防止标签泄露。患者的其余表型特征用于训练一个简单的自动编码器, 以通过解码器重建输入, 并通过分类器使用潜在的低维数据表示(编码器的输出)对ASD状态进行分类。使用**均方误差的加权和**作为无监督(对于标记和未标记的训练样本)和**交叉熵**作为有监督(对于已标记的训练样品)损失函数来训练自动编码器。

Table 2: Results corresponding to boxplot of Fig. 2 demonstrating performance on ABIDE dataset when the graph modality is not available in the testing phase

Method	Accuracy	F1 Macro	AUC
Abraham et. al	0.607 \pm 0.0	0.593 \pm 0.0000	0.656 \pm 0.0
DNN	0.629 \pm 0.0065	0.616 \pm 0.0099	0.651 \pm 0.0086
DNN-JFC	0.632 \pm 0.0185	0.628 \pm 0.0119	0.663 \pm 0.0121
Parisot et. al	0.6 \pm 0.0130	0.578 \pm 0.0076	0.667 \pm 0.0103
EV-GCN	0.627 \pm 0.0092	0.6210 \pm 0.0121	0.663 \pm 0.0074
GKD(proposed)	0.667 \pm 0.0087	0.662 \pm 0.0082	0.695 \pm 0.0039

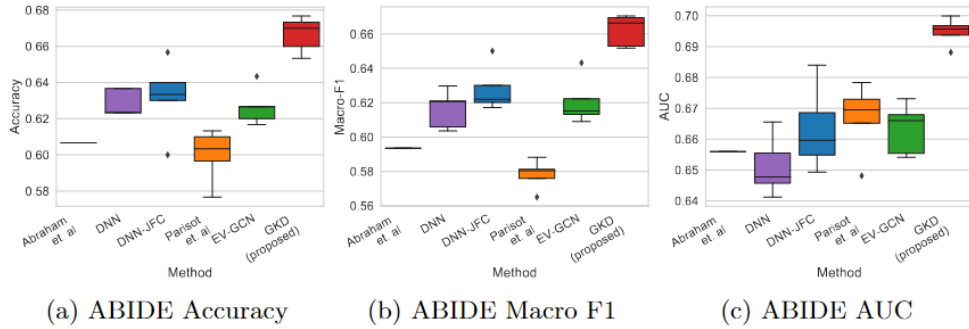


Fig. 2: Boxplot results of compared methods on the ABIDE dataset

2. 阿尔茨海默病神经成像计划 (ADNI)

Table 3: Results on TADPOLE dataset when the graph is available in the testing phase and testing samples are also accessible during the training

Method	Accuracy	F1 Macro	AUC
Abraham et. al	0.607 \pm 0.0	0.593 \pm 0.0	0.656 \pm 0.0
DNN	0.629 \pm 0.0065	0.616 \pm 0.0099	0.651 \pm 0.0086
DNN-JFC	0.65 \pm 0.0152	0.641 \pm 0.0197	0.688 \pm 0.0067
Parisot et. al	0.704 \pm 0.0252	0.685 \pm 0.0398	0.748 \pm 0.0106
EV-GCN	0.701 \pm 0.0309	0.69 \pm 0.028	0.767 \pm 0.0216
GKD(proposed)	0.714 \pm 0.0076	0.707 \pm 0.0091	0.769 \pm 0.005

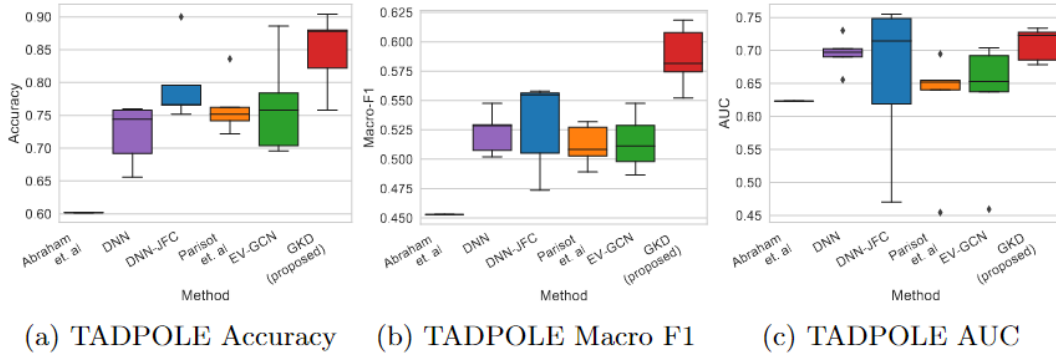


Fig. 3: Boxplot results of compared methods on the TADPOLE dataset

3. 合成数据集

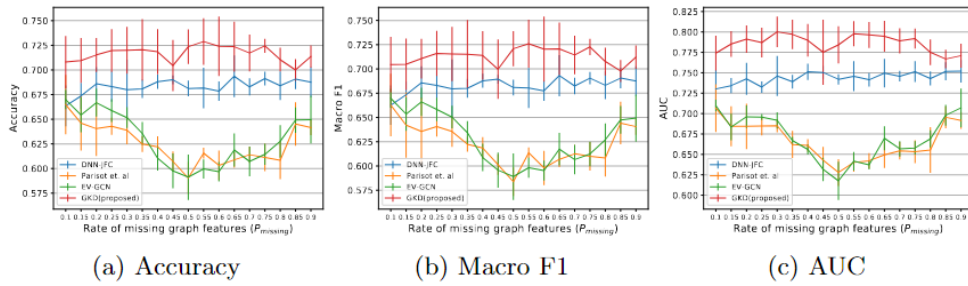


Fig. 4: Results of multi-modal methods by changing the rate of missing value in the graph-features at training

结论

论文中提到了GKD是一种新的半监督方法，旨在保留深度网络学习到的信息，但将这些知识传递到一个更小、更高效的网络上，该网络仅使用非图形模态来操作。GKD方法受到知识蒸馏框架的启发，通过LPA算法来训练具有神经网络结构的“Teacher”模型进行半监督标签传播，同时考虑到图像模态。然后教师网络使用所有的模态和非图像数据生成软伪标签，这些标签捕捉输入样本之间的相似性。该方法通过图形获得了标签的知识传递，可以有效地将知识在样本或节点之间进行转移。接下来，使用伪标签和非图像数据仅训练学生神经网络，该网络的参数要比教师网络少，适合在资源有限的设备上部署。论文在公开的两个数据集，自闭症谱系障碍 (ABIDE) 和阿尔茨海默病神经成像计划 (ADNI) 上对GKD方法进行了评估，并在内部多模态数据集上进行了彻底的分析。实验结果表明，GKD方法在准确性、曲线下面积 (AUC) 和宏平均 F1 得分等方面优于现有的基准方法。同时，GKD方法表现出对不同模态的鲁棒性，即使在测试数据中缺失了图像模态也可以使用。此外，该方法为未来的研究提供了一个框架，旨在利用各种模态中的隐藏结构，学习具有卓越性能的多模态模型。