

【动手做】 使用图变分自编码 VGAE完成链接预测

讨论内容

0 工具准备

- Python和Anaconda
- DGL-Deep Graph Library

1 知识准备

- 自编码器Auto-Encoder
- 图变分自编码VGAE

2 案例演示

- 代码运行流程
- 代码演示

3 参考资料

4 回顾 【动手做】 图神经网络模型

AchieveFun

1 知识准备：自编码器Auto-Encoder

- 参考幻灯片 <https://github.com/bettermorn/IAICourse/blob/main/courseware/%E5%B7%A5%E4%B8%9A%E6%99%BA%E8%83%BD%E5%AE%9E%E6%88%987-%E7%94%9F%E4%BA%A7%E7%AE%A1%E7%90%86%E4%B8%8E%E5%BC%82%E5%B8%B8%E8%AF%86%E5%88%AB.pdf>
- 视频： <https://www.bilibili.com/video/BV16o4y1L7uF> **【专题7:生产管理之异常检测】**
2.Bert、GPT、AutoEncoder原理介绍 50分左右开始
- 视频： <https://www.bilibili.com/video/BV1KM4y1h7Pt> **【专题7:生产管理之异常检测】** 实验
展示：AutoEncoder Detection
- 变分自编码器VAE：学到低维向量表示的分布（通常为高斯分布）。在低维向量表示的分布中采样得到低维向量表示，接下来经过解码器还原出原始样本。

1 知识准备：图变分自编码VGAE

- 输入：节点特征矩阵 \mathbf{X} 和图结构的邻接矩阵 \mathbf{A} 。
- 编码器：推理模型，两层GCN，学习潜在分布的 $q(\mathbf{Z}|\mathbf{X},\mathbf{A})$ 的均值和标准差。
- 解码器：生成模型，潜在变量之间的内积。相似度量。
- 目标：学习数据的生成分布。
- 通过预测图结构，即邻接矩阵中的0和1，构造一个无监督的重建任务。待训练收敛后，编码器可用来提取图节点表示。

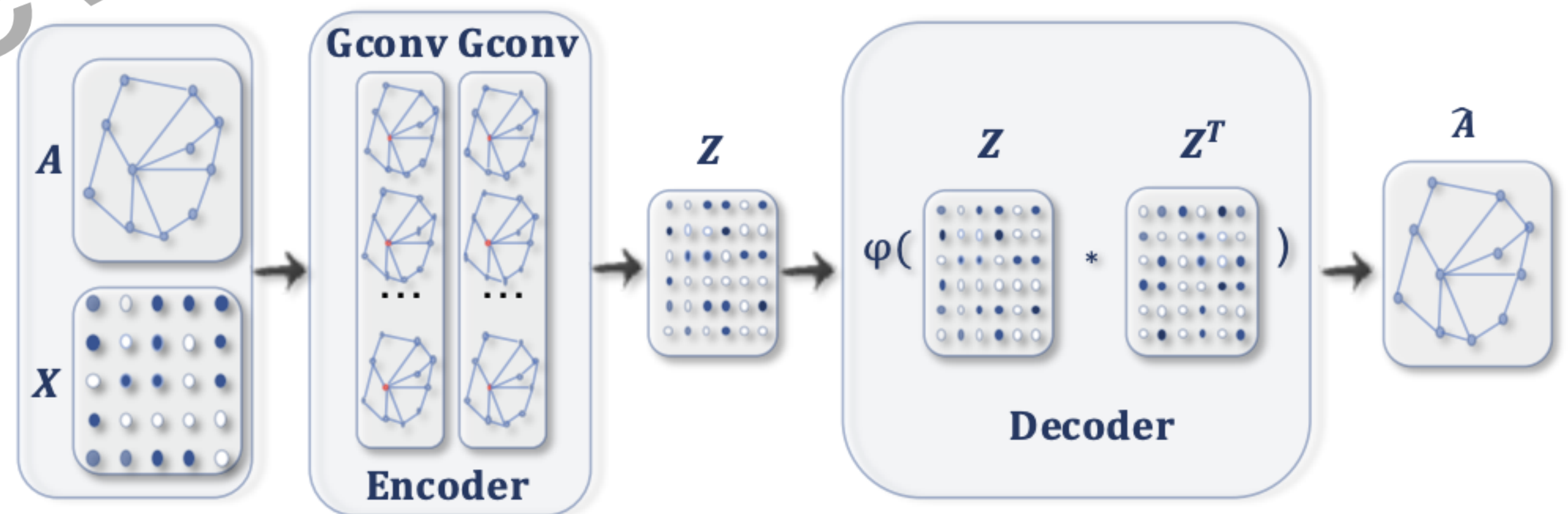
Inference model We take a simple inference model parameterized by a two-layer GCN:

$$q(\mathbf{Z}|\mathbf{X},\mathbf{A}) = \prod_{i=1}^N q(\mathbf{z}_i|\mathbf{X},\mathbf{A}), \text{ with } q(\mathbf{z}_i|\mathbf{X},\mathbf{A}) = \mathcal{N}(\mathbf{z}_i|\boldsymbol{\mu}_i, \text{diag}(\boldsymbol{\sigma}_i^2)). \quad (1)$$

Generative model Our generative model is given by an inner product between latent variables:

$$p(\mathbf{A}|\mathbf{Z}) = \prod_{i=1}^N \prod_{j=1}^N p(A_{ij}|\mathbf{z}_i,\mathbf{z}_j), \text{ with } p(A_{ij}=1|\mathbf{z}_i,\mathbf{z}_j) = \sigma(\mathbf{z}_i^\top \mathbf{z}_j), \quad (2)$$

图片来源：<https://arxiv.org/pdf/1611.07308.pdf>



图片来源：<https://arxiv.org/pdf/1901.00596.pdf> A Comprehensive Survey on Graph Neural Networks

Variational Graph Auto-Encoders

图变分自编码VGAE：自然地纳入节点特征

- We introduce the variational graph auto-encoder (VGAE), a framework for unsupervised learning on graph-structured data based on the variational auto-encoder (VAE). This model makes use of latent variables and is capable of learning interpretable latent representations for undirected graphs. We demonstrate this model using a graph convolutional network (GCN) encoder and a simple inner product decoder. Our model achieves competitive results on a link prediction task in citation networks. In contrast to most existing models for unsupervised learning on graph-structured data and link prediction, our model can naturally incorporate node features, which significantly improves predictive performance on a number of benchmark datasets.

- 来源: <https://arxiv.org/pdf/1611.07308.pdf>

机器翻译:

我们介绍了变异图自动编码器 (VGAE)，这是一种基于变异自动编码器 (VAE) 的图结构数据无监督学习框架。该模型利用潜变量，能够学习无向图的可解释潜表征。我们使用图卷积网络 (GCN) 编码器和简单的内积解码器演示了这一模型。我们的模型在引文网络的链接预测任务中取得了有竞争力的结果。与大多数用于图结构数据无监督学习和链接预测的现有模型相比，我们的模型可以自然地纳入节点特征，从而显著提高了在一些基准数据集上的预测性能。

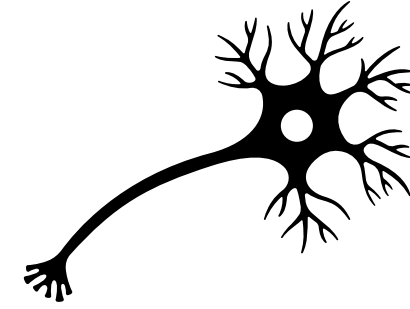
2 案例演示

- 案例来源：
 - <https://paperswithcode.com/paper/variational-graph-auto-encoders>
- 代码
 - <https://github.com/dmlc/dgl/tree/master/examples/pytorch/vgae>

学习目标

- 编码器：两层图卷积网络
- 解码器：两两计算两点之间存在边的概率来重构图
- 损失函数：生成图和原始图之间的距离度量，以及节点表示向量分布和正态分布的散度
- 来源： <https://github.com/dmlc/dgl/tree/master/examples/pytorch/vgae>

环境准备



1. 安装Anaconda <https://www.anaconda.com/download/>
2. 在Anaconda建立环境dgl, 根据dgl文档(<https://www.dgl.ai/pages/start.html>)确定**Python版本**
3. 根据代码, 使用pip install -r requirements.txt 安装依赖库并验证库的存在pip list
 - o torch
4. 安装dgl <https://www.dgl.ai/pages/start.html>

例如: conda install -c dgteam dgl

代码结构

文件名	作用	方法
input_data.py	将数据转换为邻接矩阵	parse_index_file load_data
preprocess.py	预处理图	mask_test_edges ,mask_test_edges_dgl sparse_to_tuple ,preprocess_graph
model.py	VGAEModel	__init__ encoder decoder forward
train.py	训练模型	compute_loss_para,get_acc, get_scores,dgl_main, web_main



2.1 代码运行流程

- 问题：在一个图（如引文网络）中对节点（如文档）进行分类，在这个图中，只有一小部分节点有标签。
 - Cora数据集由 2708 篇科学出版物组成，分为七类。引用网络由 5429 个链接组成。数据集中的每份出版物都由一个 0/1 值的单词向量来描述，表示字典中相应单词的缺失/存在。字典由 1433 个独特的单词组成。
- 训练流程：
 1. 装载数据集
 2. 抽取节点特征
 3. 生成输入
 4. 用 10%的正向链接建立测试集
 5. 创建训练图
 6. 创建模型
 7. 评估模型

2.2 代码演示

- <https://github.com/dmlc/dgl/tree/master/examples/pytorch/vgae>
- <https://arxiv.org/pdf/1611.07308.pdf>

AchieveFun

3 参考资料

- <https://arxiv.org/pdf/1611.07308.pdf>
- <https://github.com/dmlc/dgl/tree/master/examples/pytorch/vgae>
- <https://arxiv.org/pdf/1901.00596.pdf>

AchieveFun

4 回顾【动手做】图神经网络模型

- GraphSAGE: 链接预测
- GCN图卷积网络: 分类
- GAT图注意力网络: 分类
- VGAE 图变分自编码模型: 链接预测
- DiffPool: Hierarchical Graph Representation Learning with Differentiable Pooling 利用可变池进行层次图表示学习
- GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training 图对比学习模型 自监督学习
- MVGRL: Contrastive Multi-View Representation Learning on Graphs 图上的对比多视图表征学习
- HGT: Heterogeneous Graph Transformer 异构图转换器
- Text Generation from Knowledge Graphs with Graph Transformers 利用图转换器从知识图谱生成文本
- GPT-GNN: Generative Pre-Training of Graph Neural Networks 生成式图网络预训练框架

总结

0 工具准备

- Python和Anaconda
- DGL-Deep Graph Library

1 知识准备

- 自编码器Auto-Encoder
- 图变分自编码VGAE

2 案例演示

- 代码运行流程
- 代码演示

3 参考资料

4 回顾 【动手做】 图神经网络模型

AchieveFun