**基于序列到序列模型的图匹配问题**

一、问题描述

给定图查询图 ,和目标图 ，假如查询图 与目标图 的子图存在匹配关系，那么可以找到一种映射关系且满足以下条件：

1. 对于任意的一点 都存在满足,
2. 对于任意的一边，都存在边

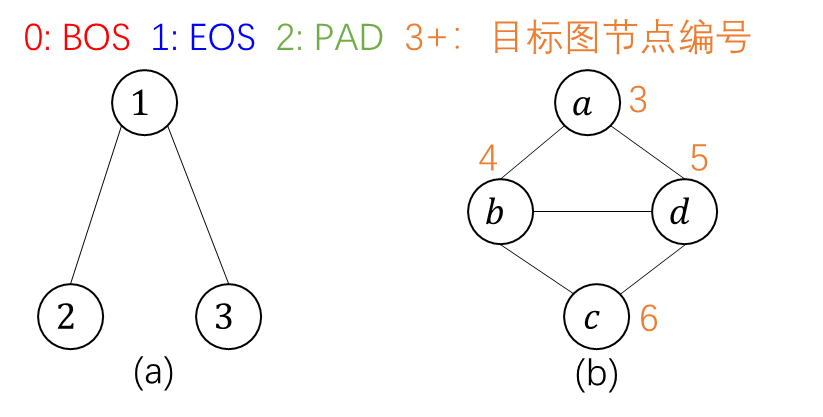


图1.（a）查询图q；（b）目标图t。

如上图所示，查询图1(a)和1(b)至少存在两种匹配关系。

1. 1-a;2-b;3-d
2. 1-b;2-a;3-c

二、参考阅读

[1] Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks.In Advances in Neural Information Processing Systems (NIPS 2014).

[2] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In Advances in Neural Information Processing Systems, pp. 2692–2700, 2015.

[3] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. In International Conference on Learning Representations, 2016.

[4] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How POWERFUL ARE GRAPH NEURAL NETWORKS? arXiv:1810.00826v3 [cs.LG] 22 Feb 2019

[5] Kipf, T. N., and Welling, M. 2017.Semi-supervised classification with graph convolutional net-works. In ICLR

[6] Liang Yao, Chengsheng Mao, Yuan Luo. Graph Convolutional Networks for Text Classification. arXiv:1809.05679v3 [cs.CL] 13 Nov 2018

三、数据集合

从200个点查询图T中(target.npy)，随机游走产生10000个包含20-40个点的查询图集合Q，Q中的每个查询图及其在目标图T中的位置编号构成一个样本。10000个样本中，其中80%用作训练集和，20%用作测试集合。如图1所示的两种匹配方式，数据样本存储格式如下：

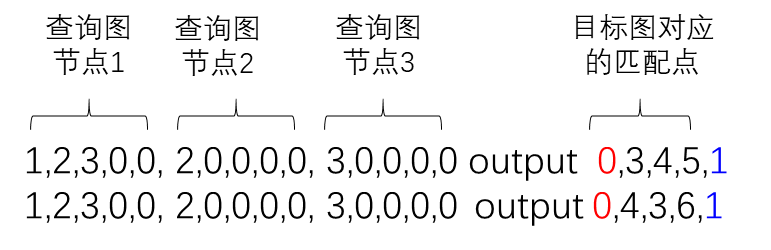


图2 数据样本格式

其中，查询图节点编码如第四节描述。

**注意：本项目所针对的目标图T是保持不变的，给定一个新的查询图（20-40节点）的时候预测其在T中的匹配节点位置。**

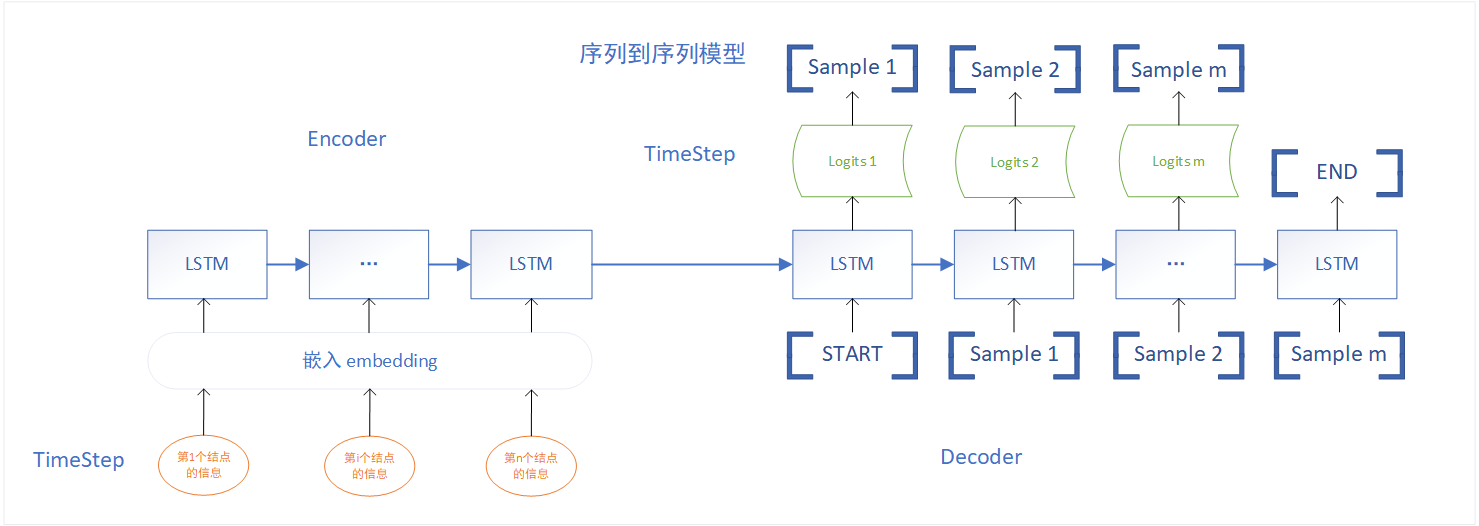
四、建议方法

**1. 查询图节点序列编码**

表1 查询图节点编码格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 节点编号 | 子节点1 | 子节点2 | 子节点3 | … | 子节点N |
| 1 | 2 | 3 | 0 |  | 0 |
| 2 | 0 | 0 | 0 |  | 0 |
| 3 | 0 | 0 | 0 |  | 0 |

**2.神经网络的结构：**



采用序列到序列模型，进行查询图在目标图中的目标点预测。

1. **可能优化的方向:**

（a）采用更好的图节点编码方式，如采用图卷积神经网络进行图节点嵌入编码，提升预测能力。

（b）在序列到序列模型中采用 attention/self-attention机制，关注全局信息，提升预测能力。

五、评估方法

1. 测试集的匹配准确率。