

DBIS 实验室 2022

面试要求说明

科研任务一：

ICLR 2017, A STRUCTURED SELF-ATTENTIVE SENTENCE EMBEDDING

任务目标：

- 讲解论文；
- 至少复现论文的 table 1 和 figure 2；

科研任务二：

ICML 2015, From Word Embeddings To Document Distances

任务目标：

- 讲解论文；
- 复现论文的 figure 3 和 figure 7（可只复现部分数据集）；

科研任务三：

ACL 2020, Graph Neural News Recommendation with Unsupervised Preference Disentanglement

任务目标：

- 讲解论文；
- 复现论文的 Table 2（可只复现部分方法）和 figure 4

科研任务四：

TKDE 2020, Dynamic Heterogeneous Information Network Embedding with Meta-path based Proximity

任务目标：

- 讲解论文；
- 实现论文中的模型 St-HNE 或 Dy-HNE（Section 4）；
- （St-HNE）复现 5.2 节的节点分类任务和关系预测任务，参考 Table 2,3；
或（Dy-HNE）复现 5.3 节的节点分类任务和关系预测任务，参考 Table 4,5。

注：

- 实验数据集可自己选择，不限于论文中的数据集；
- 论文中给出了所用对比方法源码，可从中选择几种进行对比实验。

相关资料：

- 不限定实现语言，如使用 python 进行实现可使用线性代数库 `scipy.linalg`
 - <https://docs.scipy.org/doc/scipy/reference/tutorial/linalg.html>
- 论文中给出了所用数据集的来源，可自行选择下载，或选用其他合适的数据集。

- http://www.shichuan.org/HIN_dataset.html
- <http://snap.stanford.edu/data/index.html>

上述项目包含大量可用于图实验的数据集，可自行选取数据集进行实验。

科研任务五：

SIGMOD 2014, The Pursuit of a Good Possible World: Extracting Representative Instances of Uncertain Graphs

任务目标：

- 讲解论文
- 实现论文中的 ADR 和 ABM 模型
- 复现论文中关于 degree distribution、clustering coefficient、shortest path distance、betweenness centrality 指标的实验结果，即 Figure 6 - Figure 9。在一个数据集上实现即可，四个指标可任选其二。

注：

- 实验数据集可自己选择，不限于论文中的数据集。

相关资料：

- 数据集下载网站
斯坦福大学图网络分析平台 SNAP
<http://snap.stanford.edu/>
- 若使用 C++ 开发，可使用斯坦福大学提供的 SNAP 接口；
官方网址：<http://snap.stanford.edu/snap/index.html>
若使用 Python 开发，可使用 networkx 模块实现图构建，或者使用斯坦福大学提供的 SNAP 接口（Python 版本）
官方网址：<http://snap.stanford.edu/snappy/index.html>
指导手册：<http://snap.stanford.edu/snappy/doc/reference/index-ref.html>
- 相关指标可使用现有的库函数，例如 networkx 中的 degree_histogram 函数。SNAP 接口中也提供了对应函数，请自行查阅。
- 图表生成可使用 Python 的 Matplotlib 模块。

科研任务六：

KDD 2020, AM-GCN: Adaptive Multi-channel Graph Convolutional Networks

任务目标：

- 讲解论文
- 实现论文中的模型 AM-GCN (section3)，并复现 4.2 节的节点分类任务和 4.3 节消融实验任务，参考 Table2 和 Figure2。

注：

- 实验数据集可自己选择，不限于论文中的数据集
- 论文附录中给出了所用对比方法源码，建议也选择其中一二种对比实验。

相关资料：

- 论文提出的模型用到了 k 邻近算法，若使用 Python 开发，可借助第三方库 sklearn，参见 <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- 论文模型基于图卷积网络构建，关于图卷积网络可参考以下资料：

- <https://arxiv.org/abs/1609.02907>
- <https://zhuanlan.zhihu.com/p/71200936>
- <https://jonathan-hui.medium.com/graph-convolutional-networks-gcn-pooling-839184205692>

3. 论文附录中给出了所用数据集的来源，可自行选择下载，或选用其他合适的数据集。