

知识表示实验报告

162050128 杨铭 162050120 冉鹏
162050127 颜劭铭 162050121 赵家琛

事理知识图谱

日期：2023 年 5 月 26 日

摘 要

本实验中，我们实现了一个基于事理知识图谱的项目。我们通过自行开发的网络爬虫从凤凰新闻网和新华网获取了大量新闻标题数据。然后，使用 LTP 和 jieba 工具进行自动分词和三元组进行粗提取，然后手动微调提取效果。为了提升模型的泛化性能，我们同时使用了 Title2Event 论文 [3] 中提供的数据。我们进一步利用 word2vec 模型进行词向量训练，并将三元组知识存储于 Neo4j 图数据库中。实验结果验证了所构建的知识图谱在信息分析和推理任务中的有效性。未来，将进一步拓展和优化该知识图谱，以应对更复杂和多样化的实际应用场景。我们的代码开源在 <https://github.com/Starry-16/KRR>。

1 简介

1.1 事理知识图谱

事理知识图谱是一种基于知识图谱技术构建的特定领域知识图谱，旨在对事理、道理、原理等概念进行系统化的整理和表示。它通过将相关的事理知识进行结构化建模，形成一种图谱化的知识表示形式，以便于知识的组织、查询和推理。但是，事理知识图谱中的节点往往代表抽象的概念、道理或原理，而不是具体的实体。这些概念可能是抽象的思想、规则、原则等，需要进行更高层次的语义建模和表示。如何准确地表示抽象的概念和复杂的关系，如何进行语义的表达和关联，是构建事理知识图谱时需要面对的难题。总之，和实体知识图谱相比，事件知识图谱具有更深入、丰富、精确的语义表示能力，可广泛应用于各种知识的学习、推理和理解，同时也更具有挑战性。

1.2 工作贡献

在实验中，我们首先开发了自己的网络爬虫，针对凤凰新闻网和新华网进行数据收集。通过爬取这两个新闻网站的文章和报道，我们获得了大量的文本数据作为知识图谱的基础。这些数据包含了各种领域的信息，涵盖了政治、经济、科技、娱乐等多个方面的内容。为了更加充实训练数据，我们从 Title2Event 论文 [3] 中获取了额外的事理知识三元组作为补充。

为了更好地处理和理解这些文本数据，我们采用了来自哈工大的 LTP (Language Technology Platform) [2] 和 jieba 工具进行自动分词。这些工具能够帮助将文本分割成粗略的事件实体，为

后续的三元组提取和词向量训练提供了基础。然后，我们使用手动精调的方式生成效果更好的三元组。

为了捕捉词语之间的语义相似性和关联性，我们使用了 word2vec 模型对文本数据进行了词向量训练。这种模型能够将词语映射到高维向量空间中，使得具有相似语义的词语在向量空间中距离更近。通过训练得到的词向量，我们能够进行词语相似性计算、语义关联分析等任务，为知识图谱的推理和搜索提供支持。

我们选择了 Neo4j 图数据库作为知识图谱的存储引擎，简单有效地存储和查询三元组知识。Neo4j 提供了灵活的图形数据模型和强大的查询语言 Cypher，使得对知识图谱的存储、管理和查询变得简单而高效。我们将三元组存储为节点和边的形式，利用 Neo4j 的图形分析功能，可以方便地发现实体之间的关系、进行复杂的查询操作，实现知识图谱的应用和推理。

在知识表示部分，我们选择了清华大学开源的 openKE 工具 [5]。OpenKE 是 THUNLP 基于 TensorFlow、PyTorch 开发的用于将知识图谱嵌入到低维连续向量空间进行表示的开源框架。我们通过 openKE 很轻易地实现了 TransE 的模型实验，结果表明，我们的模型对知识图谱有较好的表示能力。

2 实验方法

在这部分，我们将详细介绍事理知识图谱的构造流程 (图1) 和实验细节。

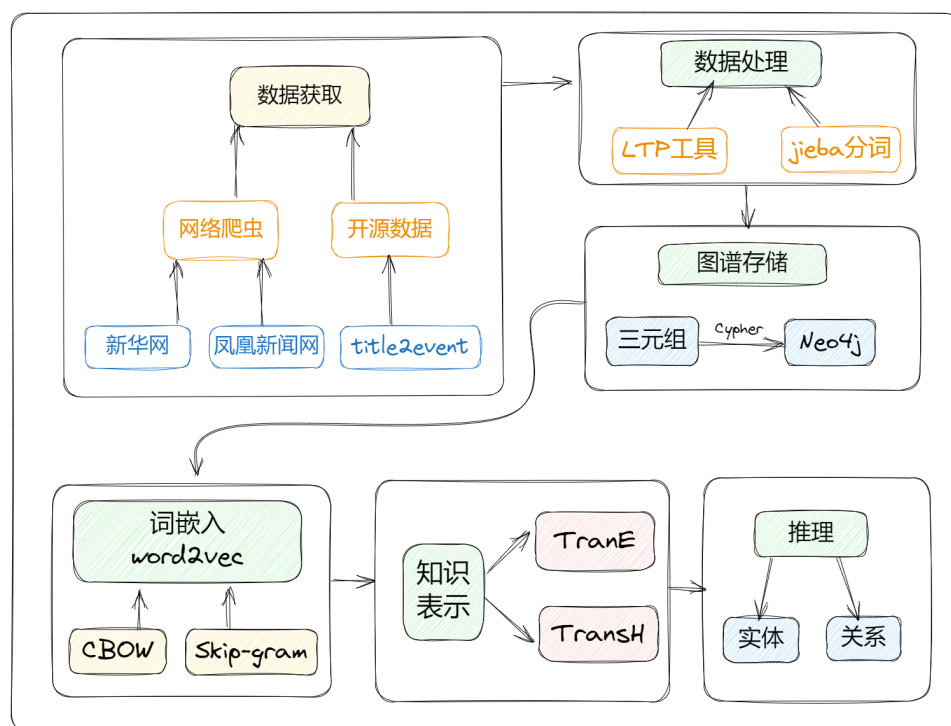


图 1: 实验流程

2.1 数据获取

我们使用 Python 编写爬虫来爬取凤凰新闻网和新华网的新闻标题获取数据。首先我们使用 requests 库发送 HTTP 请求，获取到 raw_html 之后使用 BeautifulSoup 库用于解析 HTML 页面的内容从而提取出 HTML 页面中的各个元素和数据。具体来说，使用浏览器的 F12 直接找到包含新闻标题的 HTML 元素，然后调用 BeautifulSoup 的 ‘find_all()’ 方法来查找所有匹配的 HTML 元素，最终用 ‘get_text()’ 方法来获取元素的文本内容，即新闻标题。如图2

搭建新平台 开辟新前景——多国人士积极评价中国—中亚峰会取得丰硕成果,<http://world.people.com.cn/n1/2023/05/19/c1001-32690601.html>

长安复携手 丝路启新程——国务委员兼外长秦刚就中国—中亚峰会成果接受媒体采访,<http://world.people.com.cn/n1/2023/05/19/c1001-32690602.html>

织密经贸往来 共谋繁荣发展——委员会—中亚国家经贸合作开启新篇章,<http://world.people.com.cn/n1/2023/05/20/c1002-32690603.html>

习近平向第32届阿拉伯国家联盟首脑理事会会议致贺信,<http://politics.people.com.cn/n1/2023/05/20/c1024-32690602.html>

习近平:阿盟长期致力推动阿拉伯世界联合自强,促进中东地区和平、稳定与发展,<http://politics.people.com.cn/n1/2023/05/20/c1024-32690602.html>

甘肃兰州:河口古镇“黄河灯展”亮灯,<http://pic.people.com.cn/n1/2023/05/19/c1016-32690578.html>

江苏南通:打造世界级船舶海工先进制造业集群,<http://pic.people.com.cn/n1/2023/05/19/c1016-32690579.html>

广西贵港:初夏时节 稻田风景如画,<http://gx.people.com.cn/n2/2023/05/19/c179464-40422193.html>

以雄安新区为窗 透视央企绿色发展新路径,<http://finance.people.com.cn/n1/2023/05/19/c1004-32690383.html>

新疆托克逊:万亩杏园采摘收购忙,<http://xj.people.com.cn/n2/2023/05/19/c186332-40423316.html>

“这些年,习近平主席种下这些友谊之树”,<http://politics.people.com.cn/n1/2023/05/19/c1001-32690558.html>

李强主持召开国务院常务会议,<http://politics.people.com.cn/n1/2023/05/20/c1024-32690610.html>

纪念中国伊斯兰教协会成立70周年座谈会举行,<http://politics.people.com.cn/n1/2023/05/20/c1024-32690612.html>

李希:为纪检监察工作高质量发展注入强大动力,<http://politics.people.com.cn/n1/2023/05/20/c1024-32690613.html>

“雄安新区容东片区的‘安居’和‘乐业’”,<http://www.rmxbg.cn/n2/2023/05/19/c383557-40422875.html>

辽宁推动主题教育扎实开展 凝聚振兴力量,<http://ztjy.people.com.cn/n1/2023/05/20/c457340-32690632.html>

沪上超日本 中国汽车工业换道领先,<http://finance.people.com.cn/n1/2023/05/20/c1004-32690634.html>

数读中国|五年来我国医保事业实现历史性跨越,<http://health.people.com.cn/n1/2023/05/19/c14739-32690069.html>

“贵州:万桥越万山”“黔”路不再难,<http://society.people.com.cn/n1/2023/05/20/c1008-32690625.html>

学习宣传贯彻党的二十大精神,<http://cpc.people.com.cn/GB/67481/448544/index.html>

刘国中出席2023脱贫地区农副产品产销对接会开幕式,<http://politics.people.com.cn/n1/2023/05/20/c1001-32690606.html>

李鸿忠:为全面建设社会主义现代化国家提供有力法治保障,<http://politics.people.com.cn/n1/2023/05/20/c1001-32690607.html>

何立峰出席中国—中亚委员会成立大会并讲话,<http://politics.people.com.cn/n1/2023/05/20/c1001-32690608.html>

陈文清:全面贯彻落实习近平法治思想提升法学理论研究能力和水平,<http://politics.people.com.cn/n1/2023/05/20/c1001-32690609.html>

秦刚出席哈萨克斯坦独立30周年招待会并发表讲话,<http://world.people.com.cn/n1/2023/05/20/c1002-32690605.html>

图 2: 爬虫数据

2.2 知识抽取

知识抽取任务中主要是进行命名实体识别（NER），将非结构化或半结构化的文本转化为结构化的知识表示形式，使得计算机可以利用这些知识。常用的知识抽取方法包括：

1. 基于规则的方法：使用人工定义的规则和模式来匹配和抽取文本中的特定信息。这种方法通常需要领域专家的知识 and 大量的规则工程。
2. 基于统计的方法：通过机器学习和统计技术，从标注好的训练数据中学习抽取模式，然后应用这些模式来抽取新的知识。常用的统计方法包括条件随机场（CRF）、最大熵模型等。
3. 基于深度学习的方法：利用深度神经网络模型，如卷积神经网络（CNN）和循环神经网络（RNN），对文本进行建模和抽取。这些模型可以学习到丰富的语义表示，并在大规模数据上取得良好的性能。

在本次实验中，我们由于标注数据有限，使用深度学习得到的知识抽取模型效果并不好，所以我们直接使用了哈工大的 **LTP** 工具来分析新闻标题里的事件，然后人工精细调整部分抽取效果不好的部分，最终效果如图3:



图 3: 数据清洗

2.3 实体嵌入向量计算

在这部分，我们使用 word2vec 计算实体的嵌入向量，一方面是为了计算不同实体直接的相似度，另一方面也可以作为 TransE 的输入数据，提升 TransE 的效果。

Word2Vec 是一种用于将词语映射到连续向量空间的技术，它是由 Tomas Mikolov 等人于 2013 年提出的。它基于分布式假设，即相似上下文中的词语具有相似的语义含义。

该方法一共有两种不同的实现模型：连续词袋模型（Continuous Bag of Words，简称 CBOW）和 Skip-gram 模型。这两种模型的训练方法都是基于神经网络的方法：

1. CBOW 模型：

- CBOW 模型的目标是在给定上下文词语的情况下，预测目标词语。它试图通过上下文词语的平均向量来预测目标词语。
- CBOW 模型适合于训练速度较快，对于频繁出现的词语效果较好。

2. Skip-gram 模型：

- Skip-gram 模型则相反，它的目标在给定目标词语情况下，预测上下文词语。
- Skip-gram 模型适合于处理罕见词语和具有复杂语义关系的词语。

无论是 CBOW 还是 Skip-gram 模型，它们都通过训练神经网络来学习词语的向量表示。训练过程中，模型会根据语料库中的词语上下文关系来更新词向量，以使得相似上下文的词语在向量空间中更加接近。

在实验中，由于事理知识更加抽象和复杂，所以我们主要使用 Skip-gram 来训练词向量表示，最终我们得到所有实体的高维向量表示。在图4中，简单显示了部分实体的 top 6 个相近实体。

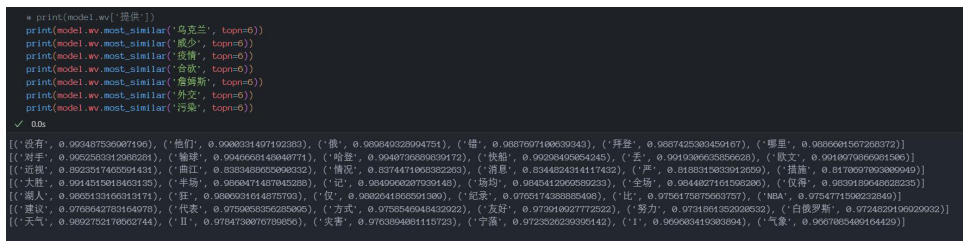
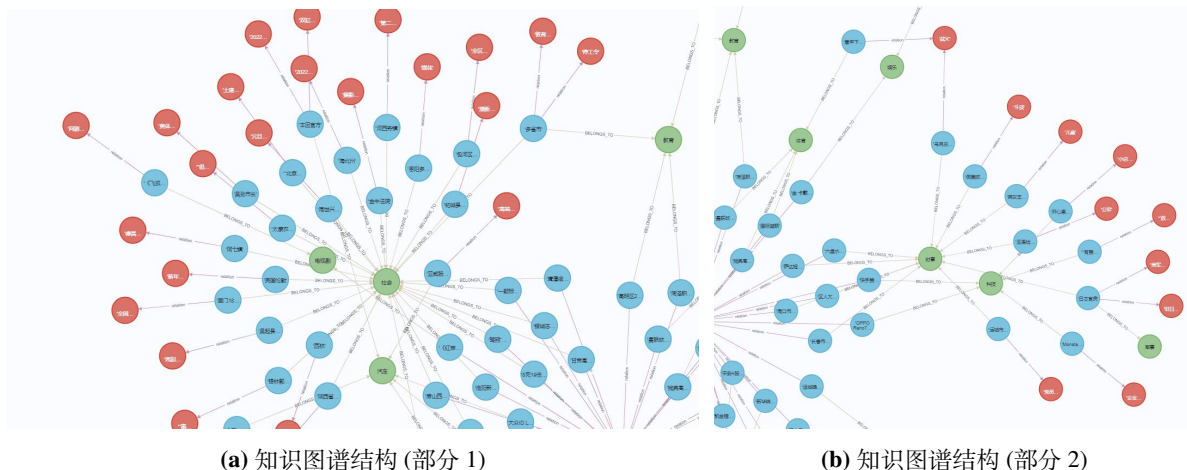


图 4: 部分词向量训练效果

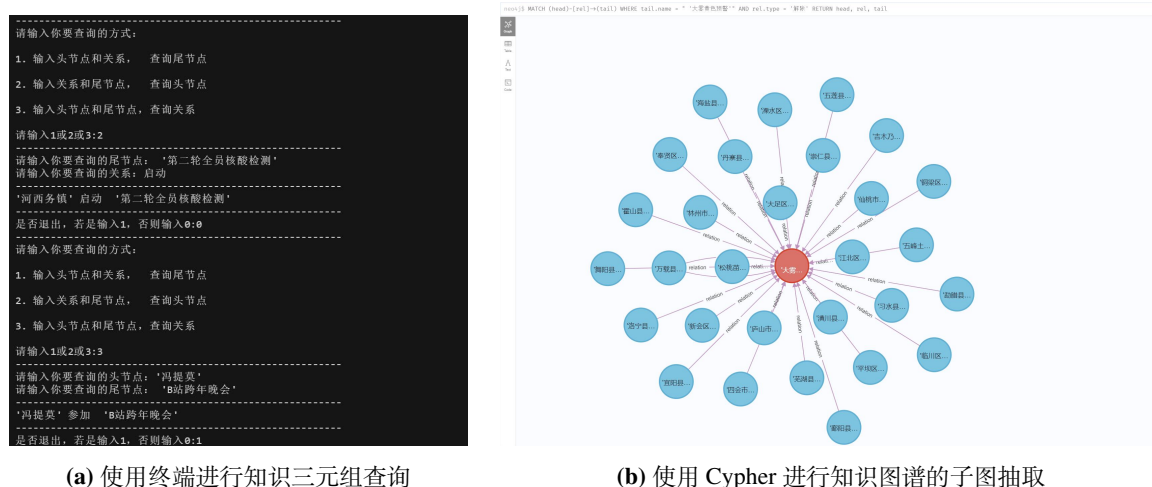
2.4 知识图谱结构

在得到精调之后的事理三元组之后，我们将其存入 Neo4j 并进行展示。首先我们定义了图数据库中的模式，即节点标签和关系类型。对于三元组存储，可以使用一个节点标签来表示实体，例如“Entity”，并使用关系类型“RELATION”表示实体之间的关系。我们通过 neo4j 的 python 接口来统一将文本的三元组 json 格式上传入 Neo4j，然后创建节点和关系。在展示环节，我们使用 Neo4j 的查询语言 Cypher 编写查询语句，以从图数据库中检索和展示三元组数据。效果如图 5a 和图 5b



2.5 知识查询

我们实验利用 Cypher 在 Neo4j 进行知识查询来验证知识存储的正确性。简单来说，如果要查询特定节点与其他实体之间的连接情况，我们使用 ‘MATCH’ 语句来查询直接的所有关系。如果我们想要查询节点的属性，则可以使用 ‘MATCH’ 语句，并在 ‘RETURN’ 子句中指定要返回的属性。我们在 CMD 终端和 Neo4j 的网页终端分别进行知识查询，效果见图 5a 和图 5b



2.6 知识表示

TransE[1] 将关系解释为在实体的低维嵌入上进行平移操作，因其在知识图谱嵌入方面具有巨大潜力而引起了广泛关注。然而，注意到 TransE 在处理关系的复杂映射属性方面表现不佳，因此基于 TransE 提出了一系列改进方法。TransH[8] 将关系建模为与一个超平面一起的平移操作。此外，TransR 和 CTransR[7] 在单独的实体空间和关系空间中构建实体和关系嵌入，取得了 SOTA 的结果。一些其他工作也试图放松 TransE 中的假设 $h + r \approx t$ ，例如将关系视为头实体向量和尾实体向量之间具有灵活幅度的平移 (TransF[4])，甚至提出了基于流形的嵌入原则 (MiniFoldE[9])，在实验证明下实现了显著改进。考虑到我们小组计算设备的限制，我们最终选择使用 TransE 和 TransH 来构建我们的实验。

2.7 TransE

TransE 是一种用于知识图谱表示学习的模型，旨在将实体和关系映射到低维连续向量空间中。它是一种基于距离的模型，用于学习实体和关系之间的语义关联。TransE 的核心思想是通过学习实体和关系的嵌入向量，使得关系三元组中的头实体向量加上关系向量近似等于尾实体向量。具体来说，TransE 假设关系三元组 (h, r, t) 中的头实体向量 h 加上关系向量 r 的结果应该与尾实体向量 t 相近。这可以通过在向量空间中最小化关系三元组的误差来实现。

对于每个关系三元组 (h, r, t) ，假设 h 、 r 和 t 分别表示为低维向量 h_i 、 r_i 和 t_i 。TransE 使用 L_1 或 L_2 范数来度量向量之间的距离，以衡量头实体向量与关系向量之和与尾实体向量之间的差异。

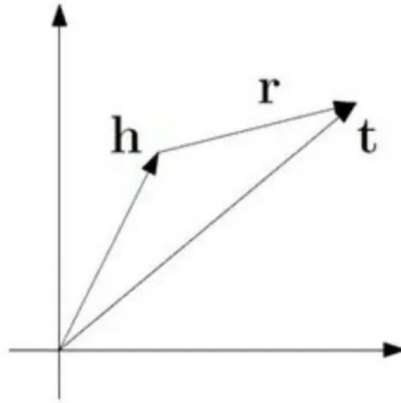


图 5: TransE 原理简单示例

在 TransE 中，关系三元组的目标是最小化以下损失函数：

$$L = \sum (||h_i + r_i - t_i||)$$

通过最小化损失函数，TransE 的目标是将正样本的距离最小化，同时将负样本的距离最大化。这样，模型可以在向量空间中学习到实体和关系之间的语义关联。

在训练过程中，可以使用随机梯度下降等优化算法来最小化损失函数，并更新实体和关系的嵌入向量。训练完成后，可以使用学习到的嵌入向量来进行实体和关系的推理和查询。

总结一下，TransE 是一种基于距离的知识图谱表示学习模型，通过学习实体和关系的嵌入

向量，使得关系三元组中的头实体向量加上关系向量近似等于尾实体向量。通过最小化损失函数，TransE 能够在向量空间中学习到实体和关系之间的语义关联。

2.8 TransH

TransH 同样是一种用于知识图谱表示学习的模型。TransH 的核心思想是，每个关系都有一个关系特定的超平面，实体的嵌入向量在这个超平面上进行投影。这个超平面的选择是通过关系向量的嵌入向量来确定的。具体来说，对于每个关系 r ，TransH 引入一个关系向量 r_i 和一个超平面向量 w_i 。实体向量 e_i 通过投影操作，将其投影到关系 r 所对应的超平面上。

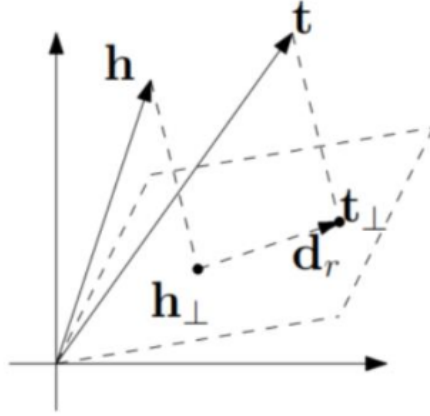


图 6: TransH 原理简单示例

为了实现这个投影操作，TransH 使用关系向量 r_i 来对实体向量 e_i 进行调整。具体来说，对于每个关系三元组 (h, r, t) ，假设头实体向量 h_i 和尾实体向量 t_i 分别表示为低维向量 h_i 和 t_i ，关系向量 r_i 表示为低维向量 r_i ，超平面向量 w_i 表示为低维向量 w_i 。投影操作可以通过以下公式来计算：

$$e_i = e_i - w_i^T e_i w_i + r_i^T e_i w_i$$

这个公式的含义是，首先用超平面向量 w_i 将实体向量 e_i 投影到超平面上，然后用关系向量 r_i 对投影后的向量进行调整，从而得到最终的实体向量 e_i 。

在训练过程中，TransH 通过最小化损失函数来学习实体和关系的嵌入向量。损失函数包括两部分：正样本的得分和负样本的得分。对于正样本，目标是使得头实体向量经过关系向量调整后与尾实体向量的距离最小化。对于负样本，目标是使得错误的头实体向量经过关系向量调整后与尾实体向量的距离最大化。

最后，通过使用随机梯度下降等优化算法，可以最小化损失函数并更新实体和关系的嵌入向量。训练完成后，可以使用学习到的嵌入向量来进行实体和关系的推理和查询。

总结一下，TransH 是一种基于超平面投影的知识图谱表示学习模型。与 TransE 相比，TransH 的主要区别在于关系特定的超平面的引入。TransE 模型中没有显式地处理不同关系之间的语义差异，它假设实体的向量表示可以通过简单的向量加法和减法来表示关系。而 TransH 通过引入超平面来建模不同关系之间的差异，更加灵活地处理了关系的语义含义。

2.9 知识推理

由于从自然语言中抽取问答信息过于复杂,加之计算资源的受限,我们无法使用大模型来增强问答能力,所以我们将推理的实验简化成 TransE/TransH 的简单实体关系的推理。根据 TransE/-TransH 的原理,我们提供三种问答方式:根据 head 和 relation 推理 tail、根据 head 和 tail 推理 relation、根据 relation 和 tail 推理。具体来说:

2.9.1 TransE 推理

1. 问句中按照先后次序包含一个实体和一个关系,如'湖人遭遇了什么?',此时对应的查询目标是尾实体,效果如图7;

```
> python query.py
Input Files Path : ./benchmarks/mydata/
The toolkit is importing datasets.
The total of relations is 12166.
The total of entities is 59103.
The total of train triples is 43967.
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
-----
请输入你要查询的方式:
1. 输入头节点和关系, 查询尾节点
2. 输入关系和尾节点, 查询头节点
3. 输入头节点和尾节点, 查询关系
-----
请输入1或2或3:1
-----
请输入你要查询的头节点: 湖人
请输入你要查询的关系: 遭遇
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100% | 1/1 [00:00<00:00, 98.03it/s]
-----
湖人 遭遇 魔鬼赛程
-----
是否退出, 若是输入1, 否则输入0:
```

图 7: TransE 推理效果 1(head+relation 推理 tail)

2. 问句中按照先后次序包含一个关系和一个实体,如'谁狂砍了 56 分?',此时对应的查询目标是头实体,效果如图8;

```
-----
请输入你要查询的方式:
1. 输入头节点和关系, 查询尾节点
2. 输入关系和尾节点, 查询头节点
3. 输入头节点和尾节点, 查询关系
-----
请输入1或2或3:2
-----
请输入你要查询的尾节点: 56分
请输入你要查询的关系: 狂砍
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100% | 1/1 [00:00<00:00, 177.41it/s]
-----
詹姆斯 狂砍 56分
-----
是否退出, 若是输入1, 否则输入0:
```

图 8: TransE 推理效果 2(tail + relation 推理 head)

3. 问句中包含两个实体,如'湖人和勇士发生了什么?',此时对应的查询目标是关系,效果如图9;

2.9.2 TransH 推理

TransH 是在 TransE 的基础上引入了关系特定的超平面,以更好地处理不同关系之间的语义差异,比如能够更好地处理多对一和一对多的关系。我们同样对于三种不同的问答类型分别做


```
-----
请输入你要查询的方式：
1. 输入头节点和关系，  查询尾节点
2. 输入关系和尾节点，  查询头节点
3. 输入头节点和尾节点， 查询关系

请输入1或2或3:3
-----
请输入你要查询的头节点：湖人
请输入你要查询的尾节点：勇士
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100%|-----| 1/1 [00:00:00:00, 248.70it/s]
-----
湖人 险胜 勇士
-----
是否退出，若是输入1，否则输入0:1
```

图 9: TransE 推理效果 3(head + tail 推理 relation)

测试，效果如下：

1. 问句中按照先后次序包含一个实体和一个关系, 如’詹姆斯担任什么角色?’，此时对应的查询目标是尾实体，效果如图10;

```
-----
请输入你要查询的方式：
1. 输入头节点和关系，  查询尾节点
2. 输入关系和尾节点，  查询头节点
3. 输入头节点和尾节点， 查询关系

请输入1或2或3:1
-----
请输入你要查询的头节点：詹姆斯
请输入你要查询的关系：担任
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100%|-----| 1/1 [00:00:00:00, 267.99it/s]
-----
詹姆斯 担任 先发中锋
-----
```

图 10: TransH 推理效果 1(head+relation 推理 tail)

2. 问句中按照先后次序包含一个关系和一个实体, 如’谁发表了 2022 年新年贺词?’，此时对应的查询目标是头实体，效果如图11;

```
-----
请输入你要查询的方式：
1. 输入头节点和关系，  查询尾节点
2. 输入关系和尾节点，  查询头节点
3. 输入头节点和尾节点， 查询关系

请输入1或2或3:2
-----
请输入你要查询的尾节点：二〇二二新年贺词
请输入你要查询的关系：发表
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100%|-----| 1/1 [00:00:00:00, 212.05it/s]
-----
澳门特别行政区行政长官贺一诚 发表 二〇二二新年贺词
-----
```

图 11: TransH 推理效果 2(tail + relation 推理 head)

3. 问句中包含两个实体, 如’南京地铁一男子对女子做了什么?’，此时对应的查询目标是关系，效果如图12;

```
-----
请输入你要查询的方式：
1. 输入头节点和关系，  查询尾节点
2. 输入关系和尾节点，  查询头节点
3. 输入头节点和尾节点， 查询关系

请输入1或2或3:3
-----
请输入你要查询的头节点：南京地铁一类子
请输入你要查询的尾节点：乘坐女子
Input Files Path : ./benchmarks/mydata/
The total of test triples is 1.
The total of valid triples is 1.
100%|-----| 1/1 [00:00:00:00, 150.69it/s]
-----
南京地铁一类子 乘坐 乘坐女子
-----
```

图 12: TransH 推理效果 3(head + tail 推理 relation)

2.10 图谱规模统计

我们存入 Neo4j 数据库的实体和关系数量如表1，从表中看出，当前知识图谱具有中等规模大小，存储的知识数量充足。此外，我们还具体统计了每个主题下的实体数量，见表2

表 1: 知识图谱中实体，关系，主题的数量统计

Head	Relation	Tail	Topic
27186	12166	33955	37

表 2: 知识图谱中具体主题下的实体数量

Topic	Count	Topic	Count
汽车	1319	音乐	159
社会	12371	文化	501
财经	6539	三农	520
房产	647	生活	118
旅游	659	科学	147
健康	344	美食	117
电影	348	综艺	412
游戏	396	搞笑	12
军事	712	情感	95
科技	2965	动漫	34
体育	5033	育儿	73
时事	4499	历史	220
天气	1013	宠物	57
教育	1451	时尚	60
电视剧	233	职场	54
娱乐	1685	曲艺	41
广告创意	3	摄影	24
星座命理	3	小说	4
壁纸头像	1		

3 结论

我们独立开发爬虫和开源数据共收集到了近 3 万实体和 1 万多关系，在 neo4j 中能够清晰展示实体和关系之间联系。通过借助 openKE 框架，我们在知识图谱上验证了 TransE 和 TransH 两种方法的性能，综合来看，都能获得较好的向量表示。为了演示推理效果，我们搭建了事理知识图谱问答系统，采用 3 种推理方式，然后人工比对分析推理正确性，都取得令人满意的效果。

3.1 项目分工

1. **162050128 杨铭**: 小组组长，主要负责知识表示模块。完成了 TransE 和 TransH 模型训练，得到了实体和关系的嵌入向量。
2. **162050127 颜劭铭**: 具体负责代码各个模块的整合、程序调试以及数据格式转换工作。在知识抽取中，负责将结构化数据中的三元组在建立 neo4j 图数据库过程中保存下来，并对基于图数据库的检索问答代码进行调试；整合各个部分的代码，使之形成完整的项目。

3. **162050121 赵家琛**: 利用已提供的基于 `ltp` 工具包的提取三元组算法对 5 个 `json` 文件中的非结构化数据进行三元组抽取并保存, 将以上两个过程中提取的三元组提供给向量表示模块负责成员。问句处理 (基于 `jieba` 分词)、自定义词典、设计正则表达式、排序的列表 (用于判断使用基于 `neo4j` 图数据库还是基于 `TransH` 模型的问答) 以及词典数据结构。
4. **162050120 冉鹏**: 主要负责项目报告的完成, 综合协调每个模块的施工进度, 疑难对接。

参考文献

- [1] Antoine Bordes et al. “Translating embeddings for modeling multi-relational data”. In: *Advances in neural information processing systems* 26 (2013).
- [2] Wanxiang Che et al. “N-LTP: An Open-source Neural Language Technology Platform for Chinese”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 42–49. doi: [10.18653/v1/2021.emnlp-demo.6](https://doi.org/10.18653/v1/2021.emnlp-demo.6). URL: <https://aclanthology.org/2021.emnlp-demo.6>.
- [3] Haolin Deng et al. “Title2Event: Benchmarking Open Event Extraction with a Large-scale Chinese Title Dataset”. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 6511–6524. URL: <https://aclanthology.org/2022.emnlp-main.437>.
- [4] Jun Feng et al. “Knowledge graph embedding by flexible translation”. In: *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*. 2016.
- [5] Xu Han et al. “OpenKE: An Open Toolkit for Knowledge Embedding”. In: *Proceedings of EMNLP*. 2018.
- [6] Qiang Li, Liwen Chen, and Yong Zeng. “The Mechanism and Effectiveness of Credit Scoring of P2P Lending Platform: Evidence from Renrendai.com”. In: *China Finance Review International* 8.3 (2018), pp. 256–274.
- [7] Yankai Lin et al. “Learning entity and relation embeddings for knowledge graph completion”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 29. 1. 2015.
- [8] Zhen Wang et al. “Knowledge graph embedding by translating on hyperplanes”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 28. 1. 2014.
- [9] Han Xiao, Minlie Huang, and Xiaoyan Zhu. “From one point to a manifold: Knowledge graph embedding for precise link prediction”. In: *arXiv preprint arXiv:1512.04792* (2015).