**Knowledge Representation Learning: A Review**

**1.基于网络形式的知识表示面临的问题:**

1）计算效率问题.基于网络的知识表示形式中,每个实体均用不同的节点表示.当利用知识库计算实体间的语义或推理关系时,往往需要人们设计专门的图算法来实现,存在可移植性差的问题.更重要的是,基于图的算法计算复杂度高、可扩展性差，当知识库达到一定规模时,就很难较好地满足实时计算的需求。

2)数据稀疏问题.与其他类型的大规模数据类似,大规模知识库也遵守长尾分布,在长尾部分的实体和关系上,面临严重的数据稀疏问题.例如,对于长尾部分的罕见实体,由于只有极少的知识或路径。涉及它们,对这些实体的语义或推理关系的计算往往准确率极低.

**2.知识表示学习的介绍**

**(1)对比优势:**

与独热表示相比,表示学习的向量维度较低,有助于提高计算效率,同时能够充分利用对象间的语义信息﹐从而有效缓解数据稀疏问题.

**(2)典型应用**

知识表示学习得到的分布式表示有以下典型应用:

1)相似度计算.利用实体的分布式表示,我们可以快速计算实体间的语义相似度﹐这对于自然语言处理和信息检索的很多任务具有重要意义.

2）知识图谱补全.构建大规模知识图谱,需要不断补充实体间的关系.利用知识表示学习模型,可以预测⒉个实体的关系﹐这一般称为知识库的链接预测(link prediction),又称为知识图谱补全(know-ledge graph completion).

3）其他应用.知识表示学习已被广泛用于关系抽取﹑自动问答、实体链指等任务﹐展现出巨大的应用潜力.随着深度学习在自然语言处理各项重要任务中得到广泛应用,这将为知识表示学习带来更广阔的应用空间.

**(3)优点**

知识表示学习实现了对实体和关系的分布式表示,它具有以下主要优点:

1)显著提升计算效率.

2)有效缓解数据稀疏.

3)实现异质信息融合.

**3.知识表示学习的方法（模型）**

**（1）距离模型**

模型评价：它对头、尾实体使用2个不同的矩阵进行投影,协同性较差,往往无法精确刻画两实体与关系之间的语义联系.

**（2）单层神经网络模型**

模型评价：虽然SLM是SE模型的改进版本,但是它的非线性操作仅提供了实体和关系之间比较微弱的联系.与此同时,却引入了更高的计算复杂度.

**（3）能量模型**

**（4）双线性模型**

**（5）张量神经网络模型**

模型评价：

由于NTN引入了张量操作,虽然能够更精确地刻画实体和关系的复杂语义联系,但是计算复杂度非常高,需要大量三元组样例才能得到充分学习.实验表明,NTN在大规模稀疏知识图谱上的效果较差.

**（6）矩阵分解模型**

可以看到RESACL的基本思想与前述LFM类似.不同之处在于,RESACL会优化张量中的所有位置,包括值为0的位置;而LFM只会优化知识库中存在的三元组.

**（7）翻译模型**

**（8）其他模型**

**4.目前存在的问题以及解决方案**

以TransE为代表的知识表示学习模型,已经在知识图谱补全、关系抽取等任务中取得了瞩目成果.但是,知识表示学习仍然面临很多挑战.这里我们以TransE为代表模型,总结认为TransE面临的3个主要挑战﹐目前已有相关工作提出一些解决方案,具体介绍如下.

存在的问题有：

**(1)复杂关系建模**

如何实现表示学习对复杂关系的建模呢?介绍了7个代表模型.

1)TransH模型

2)TransR / CTransR模型

3)TransD模型

4)TranSparse模型

5)TransA模型

6)TransG模型

7)KG2E模型

**(2)多源信息融合**

如何实现多源信息融合？

DKRL模型

文本与知识库融合的知识表示学习

**(3)关系路径建模**

为了突破TransE等模型孤立学习每个三元组的局限性,Lin等人提出考虑关系路径的表示学习方法，以TransE作为扩展基础,提出Path-basedTransE(PTransE)模型.

**5.未来研究方向**

**（1）面向不同知识类型的知识表示学习**

**（2）多源信息融合的知识表示学习**

**（3）考虑复杂推理模式的知识表示学习**

**（4）其他研究方向**