**知识图谱链接预测：**

**1.知识表示框架**

知识图谱(Knowledge Graph)是一种用图模型来描述知识和建模世界万物之间的关联关系的技术方法。知识图谱由节点和边组成。节点可以是实体，如一个人、一本书等，或是抽象的概念，如人工智能、知识图谱等。边可以是实体的属性，如姓名、书名，或是实体之间的关系，如朋友、配偶等。

随着语义网（Semantic Web）的提出，知识表示迎来了新的契机和挑战。早期Web的标准语言HTML和XML无法适应语义网对知识表示的要求，所以W3C提出了新的标准语言RDF、RDFS等。其中RDF是W3C一系列语义网标准的核心，在RDF中知识总是以三元组的形式出现。每一份知识可以被分解为头实体、尾实体和关系（或主语、宾语和谓语），表示方法如下：或

**图1：知识图谱三元组**

例如：“人工智能之父是图灵”可以写成以下RDF三元组：(人工智能，之父，图灵)。头实体，尾实体，关系。如果将三元组的头实体和尾实体看成图的节点，关系看作边，那么一个RDF知识可以被看成一个图或一个知识图谱。如下图所示，三元组是图的单元。

**2.知识图谱链接预测**

知识图谱(KGs)是真实世界信息的结构化表示。由于KGs能够以机器可读的方式对结构化、复杂的数据进行建模，因此被广泛应用于各个领域，例如问答系统、信息检索和推荐系统等。

表格

描述已自动生成

**图2：Freebase知识图谱的信息缺失**

常见的KG有FreeBase、WikiData、DBPedia、Yago等，行业KG有Google KG、Satori和Facebook Graph Search等。这些巨大的KG中包含数百万个实体和数十亿个事件。但即使是最先进的KGs也存在不完整性问题，例如FreeBase是用于研究目的的最大和最广泛使用的KGs之一，但是在FreeBase中超过70%的个体没有出生地点信息，超过99%的个体没有民族信息。这使得研究人员提出了各种各样的技术来纠正错误，并将缺失的事实添加到KGs中，通常称为知识图谱补全（Knowledge graph completion）任务。其中最常用的方法是利用KG中已经存在的事实推断缺失的事实，用以补全现有的KG，这种方法称为链接预测(Link Prediction)。

链接预测是知识图谱嵌入（Knowledge Graph Embedding）的应用之一，将知识图谱中实体和关系映射到连续向量空间中，对知识图谱中的缺失的实体或关系进行预测。例如给出头实体和关系，即，用以预测尾实体（或给出尾实体和关系，即，用以预测头实体），称为实体预测（头实体预测或尾实体预测）。给出头实体和尾实体，即，用以预测关系，称为关系预测。

表格

描述已自动生成

**图3：链接预测的三种形式**

**涉及模型：**

目前绝大多数LP模型使用原始的KG元素来学习低维表示，利用它们来推断新的事件。在短短几年的时间里，研究人员受到RESCAL和TransE等一些开创性工作的启发，开发了几十种基于不同的架构的新模型。目前主要的LP模型分为两种：

1. Translation-based module

基于翻译的模型将实体向量表示（Embedding）在低维稠密向量空间中，进行计算和推理。最早的模型之一，TransE是一个基于距离的平移模型，它将实体h、t和关系r表示为嵌入向量，并通过函数h+r=t对它们进行映射。Transe有几个扩展， TransH将关系建模为超平面；TransR将实体和关系分别嵌入到独立的向量空间中；TransD分别用两个不同的映射矩阵对头实体和尾实体进行投影。

1. Semantic-matching model

语义匹配模型利用基于相似性的评分函数。它们通过匹配实体的潜在语义和向量空间表示中包含的关系来度量事实的可信性。RESCAL模型将实体用向量表示，关系用矩阵表示。关系矩阵对潜在因素之间的成对交互作用进行了建模，使用双线性函数作为评分函数。DistMult模型将关系矩阵简化为对角矩阵，ComplEx引入复数空间对DistMult模型进行扩展，以便更好地对非对称关系进行建模。随着神经网络的兴起，出现了利用神经网络解决LP问题的模型，包括ConvE、ConvKB等。

表格

描述已自动生成

**图4：模型的损失函数、约束和空间复杂度**

**论文复现：**

1. Translation-based module

* TransE：Translating embeddingsfor modeling multi-relational data. NIPS. 2013.
* TransH：Knowledge Graph Embedding by Translating on Hyperplanes. AAAI. 2014.
* TransR：Learning Entity and Relation Embeddings for Knowledge Graph Completion. AAAI. 2015.

1. Semantic-matching model

* RESCAL：A Three-Way Model for Collective Learning on Multi-Relational Data. ICML 2011
* DistMult：Embedding entities and relations for learning and inference in knowledge bases. ICLR 2015
* ConvE：Convolutional 2D Knowledge Graph Embeddings. AAAI 2018
* ConvKB：A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network. NAACL 2018

**数据集整理：**

FB15k:

数据集情况：共有14951个mids和1345个关系类型。训练集包含483142个三元组，验证集有50000个，测试集59071。所有三元组都是唯一的，且出现在验证集和测试集中的同义词集也出现在训练集中。

下载地址：<https://download.csdn.net/download/guotong1988/10266125?utm_source=iteye_new>

FB15k-237:

数据集情况：FB15k-237是Freebase的子集，包含237种关系和14k种实体

下载地址：https://download.csdn.net/download/qq\_21097885/12340908

背景图案

描述已自动生成

WN18:

数据集情况：数据集WN18是WordNet的子集，包含18种关系和40k种实体。

下载地址： <https://download.csdn.net/download/qq_21097885/11013930>

表格

描述已自动生成

WN18RR:

数据集情况：WN18RR 是从 WN18 创建的链接预测数据集，它是 WordNet 的子集, 包含 93,003 个三元组，40,943 个实体和 11 个关系类型。

下载地址：<https://github.com/TimDettmers/ConvE/blob/master/WN18RR.tar.gz>