

知乎

首发于  
数据连接未来

# 三个角度 理解知识图谱

## 三个角度理解知识图谱



徐超

数据连接世界，智能决定未来

关注他

20 人赞了该文章

2012年谷歌首次提出“知识图谱”这个词，由此知识图谱在工业界也出现得越来越多，对于知识图谱以及相关概念的理解确实也是比较绕。自己在研究大数据独角兽Palantir之后开始接触知识图谱，也算对其有了一定了解，这里从三个角度总结一下怎么去理解知识图谱。

### 几个基本概念：本体、实体、知识库、知识图谱

- 1、本体是概念的集合，是大家都公认的概念框架，一般不会改变如“人”、“事”、“物”、“地”、“组织”，在面对对象编程里面，我们把它叫做类，在数据管理里面我们把它叫做元数据；
- 2、实体是本体、实例及关系的整合，比如“人”是本体框中的一个概念，概念中也规定了相关属性比如“性别”，小明是一个具体的人，叫做实例，所以小明也有性别，小明以及体现小明的本体概念“人”以及相关属性，叫做一个实体（简单的说就是：本体+实例）；
- 3、很多实体形成的数据库叫做知识库，如dbpedia等；
- 4、知识图谱是一种图谱组织形式，通过语义关联把各种实体关联起来，之前不叫做知识图谱，而叫做语义网。知识图谱把结构化、非结构化的数据通过数据抽取、融合在一起，体现了数据治理、语义连接的思想，有利于大规模数据的利用和迁移。



“知识图谱改变的是企业关于数据存储的传统方式，本质上就是改变数据组织的形式，利用知识的组织形式完成行业数据逻辑性的保存。”

我的理解是知识图谱的构建是基于动态本体理论的，一套完整的数据本体可以把不同的数据形式进行串联转换，融合成一个综合的数据体系去使用。如国脉数据提出的数据基因系统，我想它应该就是一套政务数据的本体框架（通俗点说就是数据目录），这套框架就像是政务数据的“基因”，其它数据都是在这个的基础之上建立的。在面对对象编程理论中，可以理解为“类”，在数据管理中，也可以理解为元数据。有了这套完整的本体框架，就为企业或政府的数据治理奠定了基础。所谓数据治理，简单的理解就是要站在战略的高度去梳理数据和业务，有什么类型的数据？怎么用？如何围绕业务构建标准的数据体系？等。而构建知识图谱的过程其实就是把各种和领域相关的数据进行一定的梳理和融合，把杂乱无章的数据通过本体框架整合成结构化的数据。浙江杭州提出的“最多跑一次”的电子政务服务理念，在具体的实施过程中也体现了这个观点。

除了有利于多源异构数据的整合，知识图谱也有利于多模（文本、视频、图片）数据的整合。最近谷歌正在研究一项工作叫做多模任务学习，就是要把文本、图片、视频多种模态一起作为输入进行学习，这是一项比较复杂的工作。但是语义连接的作用可能就是一个解决问题的方向，比如阿里的“看图说话”应用，通过深度学习识别图像中的实体和行为然后通过文本的形式呈现出来。既然图像可以转化为文本，那么视频也可以转化文本，所以通过深度学习把多种模态进行转化，利用语义进行连接和理解，这样说不定就可以解决相关问题。

## 二、语义连接角度

“深度学习学到的是事物底层特征空间，而人能理解的是事物语义空间，知识图谱能弥补这两者的鸿沟，它将人类的思维转化为机器路径思维。”

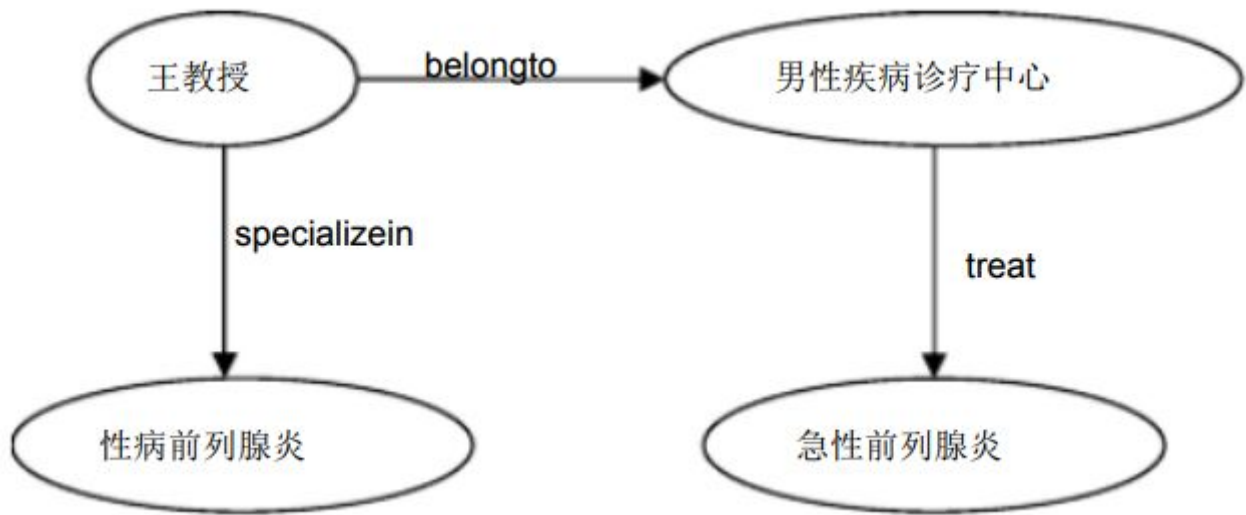
在机器学习过程中，特别让人烦的就是特征工程。就算是一个已经梳理好数据的大数据比赛，在实际的处理过程中也要考虑如何把各表的数据进行拼接，找到合适的特征或让算法自己找到合适的特征。所以如何把相关的数据关联起来，也就是如何进行数据融合，就成了使用算法的一大前提和关键。当我看到知识图谱的构建有利于进行数据融合时，就立刻吸引了我的兴趣，但后来发现，它和我们机器学习中的所体现的特征连接有点不一样。

从多源数据的融合来说，主要有如下三种方式：



2、利用特征拼接：大部分的机器学习算法都是这么做的，可以根据具体的情况增减特征数目；

3、基于语义信息的融合：通过本体之间的关系属性来体现语义信息的融合，这里面涉及到自然语言的处理、统计分析概率、语义网等。比如，“小红的爸爸（小明）”，“小红的妈妈的妈妈（小明）”这两句话中的爸爸和老公都是指的是同一个人小明，通过上下文分析、语义分析等就会获得这两人为同一个人的概率为100%。举个比较实际的知识库推理的例子，如下图



知识库中本来就有的语义信息有：<王教授， belongto ， 男性疾病诊疗中心 >以及<王 教授 ,specializein,性病前列腺炎 >。但是，知识库中并没有 <王教授， specializein, 急性前列腺 炎>，通过知识图谱的语义推理得到<王教授， specializein, 急性前列腺炎>。

这里要说明的是，深度学习的本质其实也是一种重度、高通量的“语义”链接器，只是这里的“语义”是隐藏的语义，不易理解的语义。所以有人说深度学习是一台端到端传递信息的广义翻译机器。对于知识图谱和深度学习：深度学习是通过一个黑盒子来进行预测，人类不好理解；而知识图谱是通过语义分析来推理，人类好理解。融合知识图谱与深度学习，已然成为进一步提升深度学习效果的重要思路之一。以知识图谱为代表的符号主义，和以深度学习为代表的联结主义，日益脱离原先各自独立发展的轨道，走上协同并进的新道路。目前的几种交叉应用如下：

1、利用深度学习来构建知识图谱：利用深度学习来进行实体识别、关系识别等；

2、利用知识图谱来管理数据，深度学习来做特征工程；

知乎

首发于  
数据连接未来

3、用符号表达知识图谱，利用深度学习对符号化的知识图谱进行学习将知识图谱的语义信息输入到深度学习模型中，将离散化的知识表示为连续化的向量，从而使得知识图谱的先验知识能够称为深度学习的输入；

4、利用知识作为优化目标的约束，指导深度学习模型的学习过程，通常是将知识图谱中的知识表示为优化目标的后验正则项。

### 三、智能应用角度

“知识图谱让知识可被用户访问到（搜索），可被查询（问答），可被支持行动（决策）。”

知识图谱最早就是由Google应用到知识搜索中取的一个名字，以前都不叫知识图谱。对于知识图谱的行业应用，目前主要是以一些智能问答/聊天机器人/客服类、搜索类、企业智能类、医疗服务类、图书情报服务类、金融服务类等为主，在搜索上有谷歌、百度、搜狗等，在问答上有出门问问、启齿科技的客服机器人等，在决策支持上有Palantir、明略、智器云等，当然还有一些为知识图谱提供相关组件的语义分析公司，如达观科技、智言科技等。

在问答和搜索应用上，知识图谱可以使搜索以精准的结果呈现给你，而不是返回一堆相似的页面让你自己去筛选，达到“所答即所问”，比如，搜索姚明有多高，返回来的结果就是姚明具体的身高。



知乎

首发于  
数据连接未来

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约1,130,000个

搜索工具



姚明身高：

229cm

姚明（Yao Ming），1980年9月12日出生于上海市徐汇区，祖籍江苏省苏州市吴江区震泽镇，前中国职业篮球运动员，司职中锋，现任中职联公司董事长兼总经理。1998年4月，姚明... [详情>>](#)

来自百度百科 | 报错

以前的搜索引擎返回来的基本都是已经在网页中写好的东西，比如姚明的身高，这些都是属于静态的结果。而利用知识图谱的推理功能，我们还可以获得动态的结果，在Wolfram|Alpha知识计算引擎中，我可以输入我们的数学题目，引擎可以直接给出答案，而这些答案显然并不是预先写好的。而在企业智能应用中，因为有了知识图谱，我可以关联相关的数据来探测异常、控制风险等。

series x-sinx

Input interpretation:

series  $x - \sin(x)$

Series expansion at  $x=0$ :

$$\frac{x^3}{6} - \frac{x^5}{120} + \frac{x^7}{5040} - \frac{x^9}{362880} + \frac{x^{11}}{39916800} + O(x^{12})$$

More terms

之前看到有人说，作为人工智能的产品经理，不能不了解知识图谱。为何知识图谱在人工智能时代如此重要呢？这就要从知识图谱的相关技术实现来说了。知识图谱的构建主要包括知识获取、知识建模、知识存储、知识融合、知识计算、知识表现、知识应用等。其中知识获取、知识融合是较难的点，其中主要涉及到实体识别、实体链接等，而解决这些关键的前提就是NLP技术。我们知道人工智能技术主要包括图像识别、语音识别、语义识别等，其中图像识别、语音识别技术都取得了较大的突破，而语义识别还处于起步阶段，尤其是深度学习在NLP中的应用。很多人都说智能音箱可能成为下一个入口，如果要让智能音箱操作电器等，这里面除了要让音箱知道你说了什么（利用语音识别），还要让音箱知道你说了是什么意思（利用语义识别），由此可见语义识别以及组织语义间关系的知识图谱的重要性。







首发于  
数据连接未来

编辑于 2018-01-31

知识图谱   知识管理   人工智能

▲ 赞同 20   ▼   添加评论   分享   ★ 收藏   ...

文章被以下专栏收录



数据连接未来  
一个数据农民工的思考

关注专栏

推荐阅读

知识图谱基础概念和构建

知识图谱的发展背景人工智能的发展走过了从机器智能到感知智能的阶段，正在迈向认知智能的阶段。但要实现认知智能，机器必须学会处理人类复杂的语言，学会做知识的推理，这正是人工智能遇到...

隔壁老王

知识图谱构建概述

本文是参考网络资料的学习总结，主要参考资料见文末。1.知识图谱概念知识图谱（Knowledge Graph），是结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相互关系，其基本...

朱金华



陈华钧：  
智能金融

恒生研究

还没有评论

写下你的评论...

