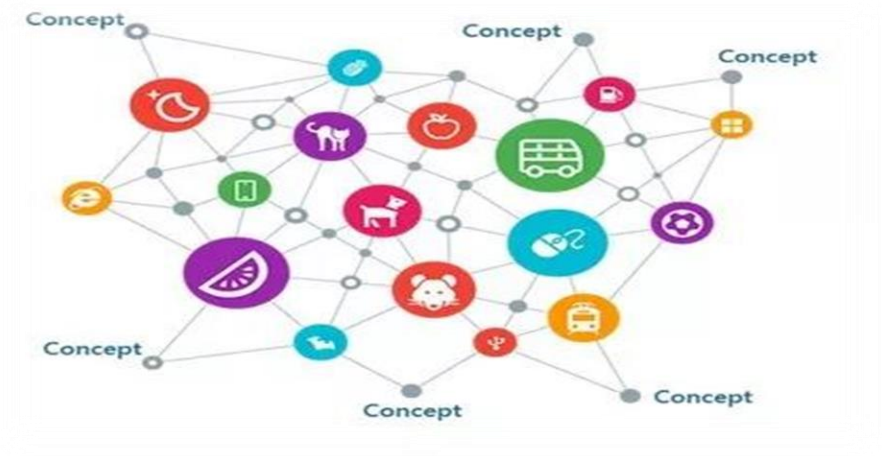
中国科学院大学
University of Chinese Academy of Sciences



知识图谱关键技术

实体--关系的识别与提取

李航航

提纲

- 背景目标
- 实体识别
- 关系提取
- 总结

一、背景目标

背景

1. 多样性

- 非结构化文本数据。
- 半结构化的网页和表格。
- 结构化数据。

2. 基础性

- 构建知识图谱的基础之一是：如何获取领域知识。
- 实体识别是从半结构化数据或非结构化数据中获取领域知识的重要方法。
- 应用：智能问答、自动摘要、信息检索、机器翻译、语义网络等。

背景目标

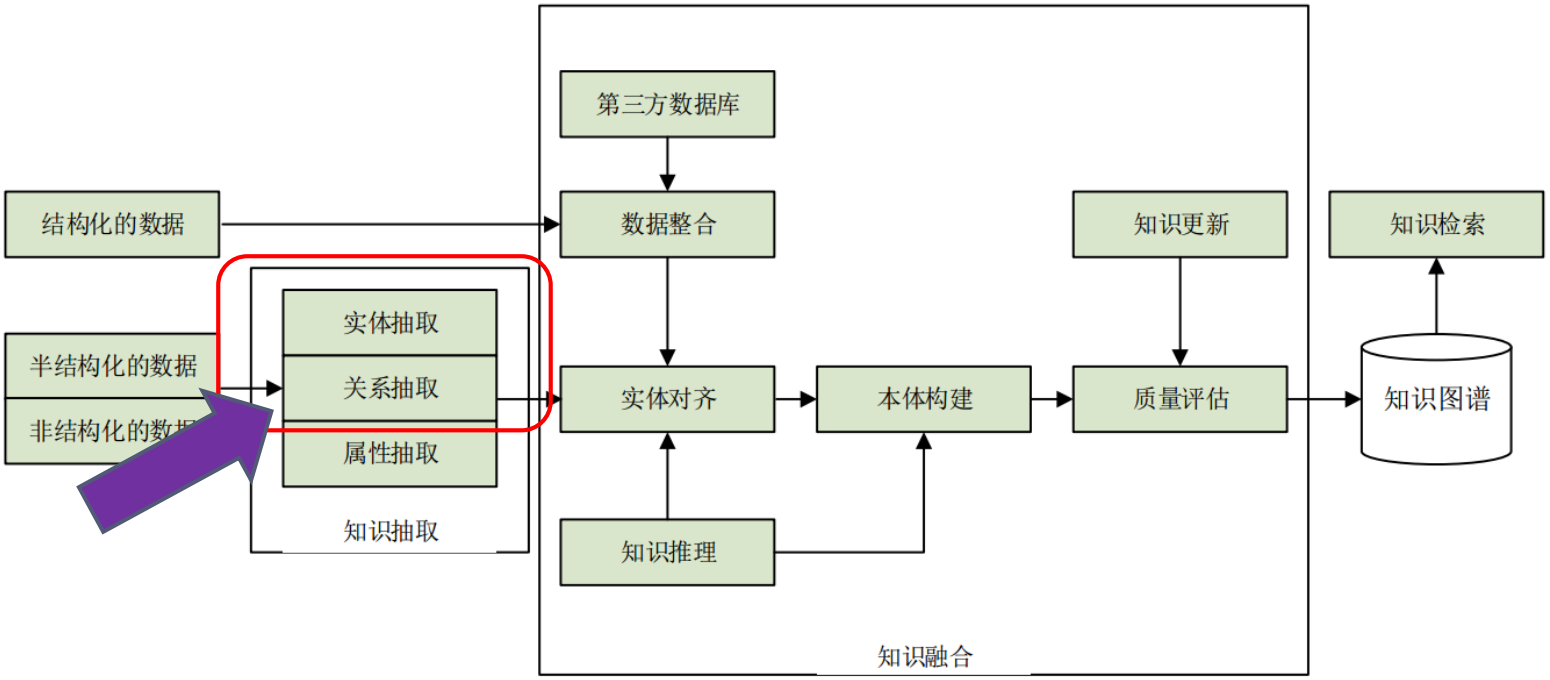
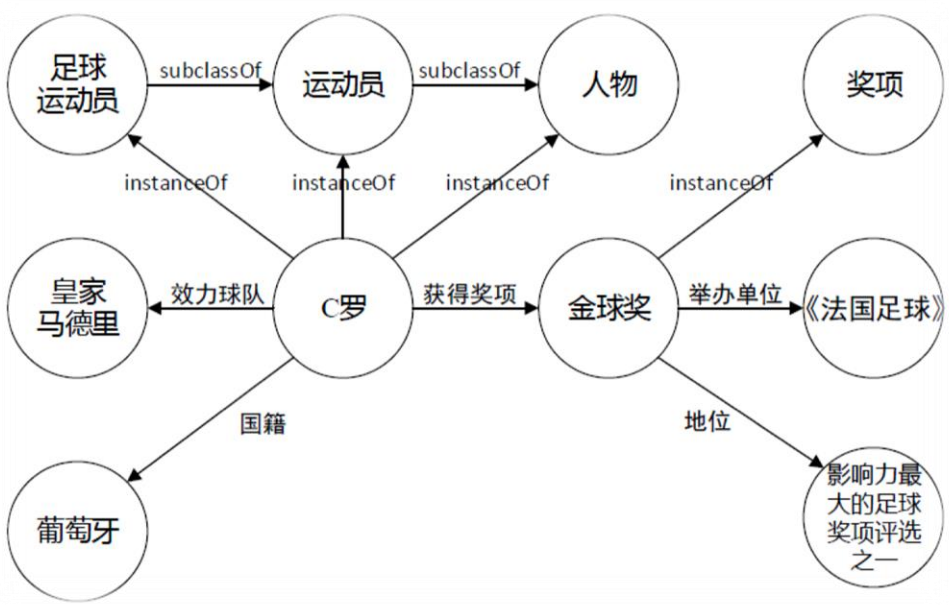


图1　知识图谱的体系架构

背景目标

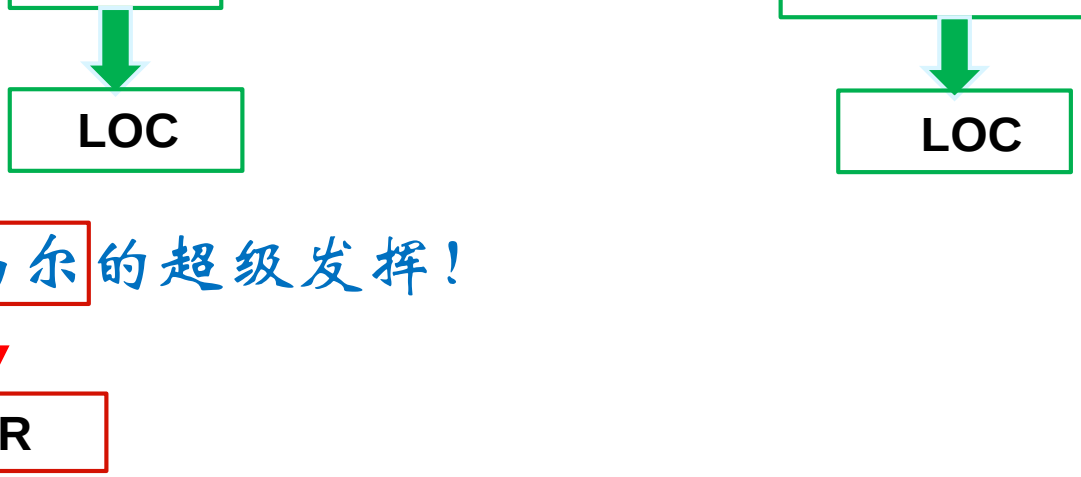
■例如：



构建“活”的知识图谱：知识自动抽取，自动生长。

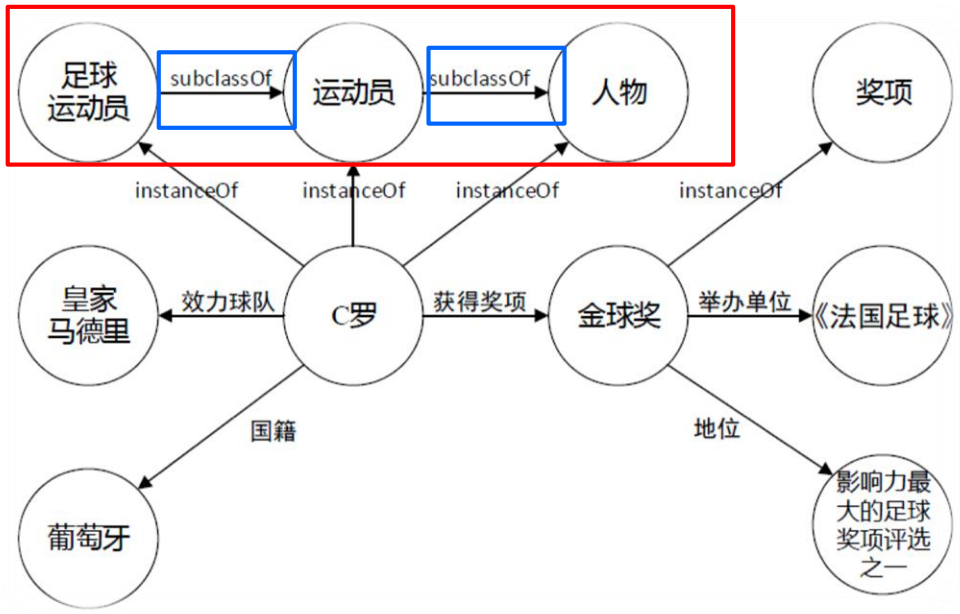
二、实体识别

- 今晚的维也纳，犹如一周之前的英国利物浦，再次见证了内马尔的超级发挥！



- 技术框架---基于特征向量的学习算法

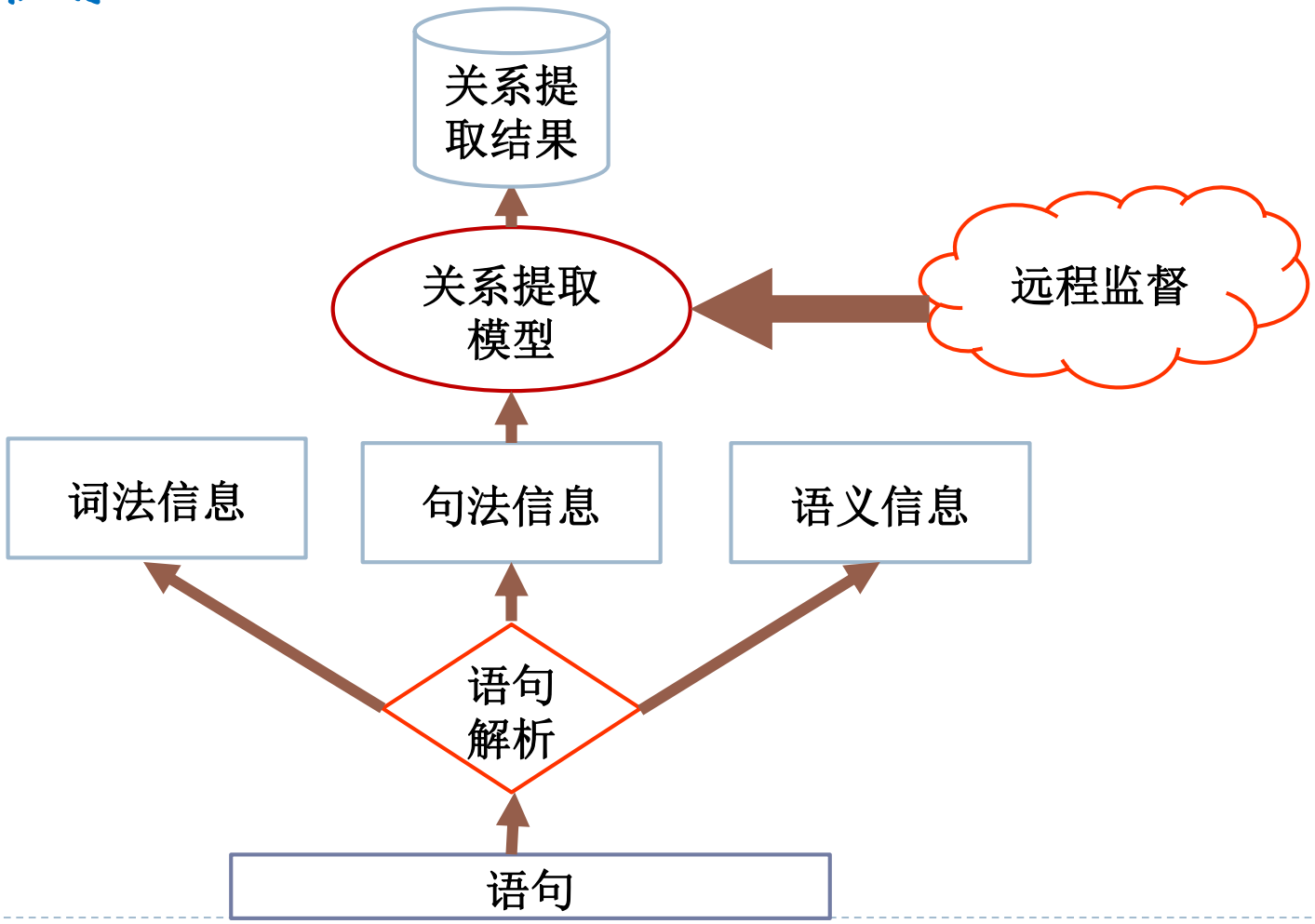
三、关系提取



知识图谱由结点和边组成，其中结点对应实体，边对应关系。

关系提取

技术框架



关系提取

◆ 语句分析

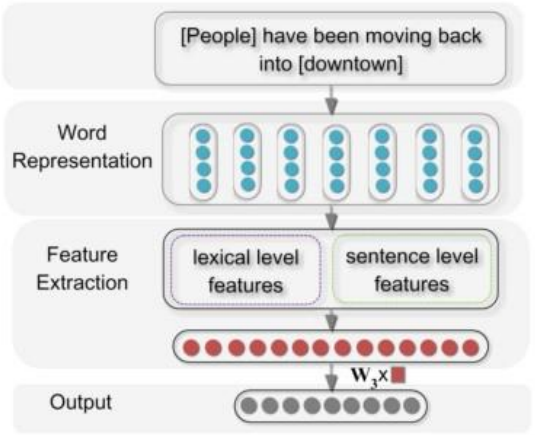
✓ 通过生成语句的句法分析树，可以获得语句的词法信息和句法信息。

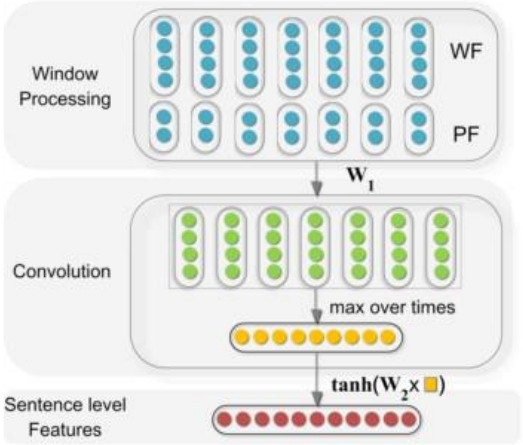
✓ 通过语句的特定结构可以获得语句的语义信息。

◆ 关系提取模型

✓ 基于CNN模型实现关系预测

包含Pooling层，以及设计了Position Features。

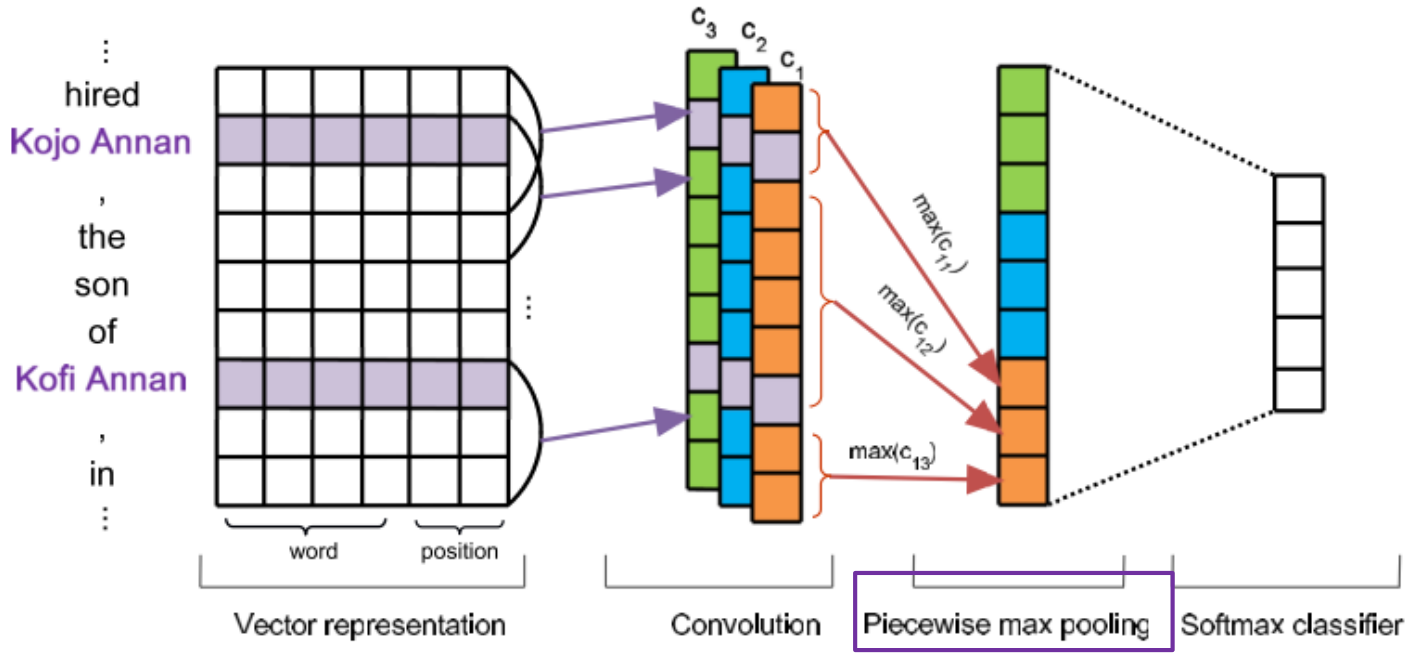




关系提取

◆ 关系提取模型

✓ 基于PCNNs模型实现关系预测



关系提取

◆ 远程监督

1. 解决有监督情况下大规模文本数据标注问题。
2. 将现有知识图谱三元组R（E1，E2）对齐到训练文本实体中，从而产生更多的训练样本。

句子	关系/分类标签	是否正确
苹果公司的创始人是乔布斯。	创始人	正确
乔布斯创立了苹果公司。	创始人	正确
乔布斯回到了苹果公司。	创始人	错误
乔布斯曾担任苹果公司的CEO。	创始人	错误

四、总结

实体识别与关系提取

- 实体识别与关系提取是构建知识图谱的重要步骤，实体识别是关系提取的前提。
- 无结构化数据量大，如何转化为结构或半结构化数据，是有效利用其数据、拓宽知识图谱使用领域的关键。
- 如何自动化进行实体识别、关系提取是增强可持续扩增能力的突破点。

总结

- 中文的命名实体识别与英文的相比，挑战更大。
- 现代汉语日新月异的发展给命名实体识别也带来了新的困难。
- 命名实体歧义严重，消歧困难。


What
Why
How