传承与变革

大模型浪潮下知识图谱研究之简单思考

任飞亮　　东北大学

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

知识图谱发展简单回顾

2012年，谷歌正式提出知识图谱概念

- 本质上，知识图谱是一种高效的**知识**组织形式

主要研究内容

- 知识图谱构建
- 知识图谱加工
- 知识图谱应用

知识图谱发展简单回顾

❖知识图谱构建

- Scheme设计
- 知识图谱基本要素抽取
  - 实体、属性、关系抽取
  - 三元组抽取
  - ……

知识图谱发展简单回顾

❖知识图谱加工

- 知识图谱补全
  - 发现当前图谱中未知的三元组事实
- 知识图谱融合
  - 实体对齐/实体链接
  - ……

知识图谱发展简单回顾

❖知识图谱应用

- 基于知识图谱的问答
- 基于知识图谱的推荐
- ……

知识图谱发展简单回顾

- 2012年，谷歌正式提出知识图谱概念
- 主要研究内容
  - 知识图谱构建
  - 知识图谱加工
  - 知识图谱应用

知识图谱基础研究任务！

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

大模型对知识图谱研究的挑战

❖大模型的特点

- **全能**：几乎可以胜任所有NLP任务
- **强大**：对大多数传统NLP任务均具有碾压式优势

❖大模型对知识图谱带来的挑战

- 知识图谱还有用吗？

大模型对知识图谱研究的挑战

❖真相是什么?

- 知识图谱构建
- 知识图谱加工
- 知识图谱应用

大模型对知识图谱研究的挑战

❖真相是什么?

- 知识图谱构建
- 知识图谱加工

结果为离散序列，并不适合大模型的应用场景，受大模型影响相对较小！

- **知识图谱应用**
  - **问答**
  - **推荐**

结果为连续序列，非常适合大模型的应用场景，受大模型影响大！

大模型对知识图谱研究的挑战

❖真相是什么?

- 知识图谱构建
- 知识图谱加工

结果为离散序列，并不适合大模型的应用场景，受大模型影响相对较小！

❖大模型的问题开始显现

- 幻觉
- 可解释性差

大模型需要知识图谱！

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

知识图谱研究中的传承与变革

- **总原则：内外兼修！**
- **内部：扩展研究宽度，提升自身竞争力**
  - 多元关系知识图谱
  - 时序关系知识图谱
  - 事理图谱
  - ……

知识图谱研究中的传承与变革

- **总原则：内外兼修！**
- **外部：与大模型相互促进**
  - 知识图谱==》大模型
    - 增强**解释性**：结果生成过程可溯源
    - 特定场景下**高可靠性**结果结果
    - **能力边界**：知道什么条件下应该说**NO！**

知识图谱研究中的传承与变革

- **总原则：内外兼修！**
- **外部：与大模型相互促进**
  - 大模型==》知识图谱
    - 利用大模型技术提升**知识图谱构建**任务的性能
      - NER、RE、RTE、EE、……
  - **东北大学TechGPT大模型属于这一方向的尝试**

TechGPT简介

- **行业属性鲜明：**
  - 可处理十几种**垂直领域行业数据**：计算机、材料、电子、机械、航空航天、金融、医药、冶金、……
- **主打知识图谱构建与智能问答**
- **文本语义理解与逻辑推理能力强**
- **ToC与ToB属性兼顾**

TechGPT简介

- **知识图谱构建类任务(ToB：LLM→KG为主)**
  - 主要针对**垂直行业**领域文本：领域术语抽取、命名实体识别、关系三元组抽取、文本关键词生成、标题生成摘要、摘要生成标题、文本领域识别、事件抽取等
- **智能问答类任务(KG→LLM为主)**
  - **ToB/ToC**：机器阅读理解式问答、基于上下文的知识问答等
  - **ToC**：建议咨询类问答、常识问答、价值观类问答等
- **序列生成任务(ToC)**
  - 文案生成、中英互译、代码生成、单轮对话、多轮对话等

侧重行业文本分析、语义理解、逻辑推理，应用前景广阔！

TechGPT实例：LLM2KG

问

请提取下面文本中的关键词。

本体是一种重要的知识库,其包含的丰富的语义信息可以为问答系统、信息检索、语义Web、信息抽取等领域的研究及相关应用提供重要的支持.因而,如何快速有效地构建本体具有非常重要的研究价值.研究者们分别从不同角度提出了大量有效地进行本体构建的方法.一般来讲,这些本体构建方法可以分为手工构建的方法和采用自动、半自动技术构建的方法.手工本体的方法往往需要本体专家参与到构建的整个过程,存在着构建成本高、效率低下、主观性强、移植不便等缺点,因而,此类方法正逐步被大量基于自动、半自动技术的本体构建方法所代替.自动、半自动构建的方法不需要（或仅需少量）人工参与,可以很方便地使用其它研究领域（如机器学习、自然语言处理等）的最新研究成果,也可以方便地使用不同数据源进行本体构建。

计算机领域

答

本体，知识库，语义信息，问答系统，信息检索，语义 Web，信息抽取，构建，自动，半自动，机器学习，自然语言处理，数据源。

TechGPT实例：LLM2KG

问 给定以下句子:"患者被诊断为慢性阻塞性肺疾病（COPD），并接受吸入性糖皮质激素治疗。"请从中抽取出领域术语。

答 该句中涉及到的领域术语包括：

1. 慢性阻塞性肺疾病 (Chronic Obstructive Pulmonary Disease, COPD) - 这是指一种慢性的、进展的肺部疾病，通常由吸烟引起，它会导致气道狭窄和炎症，导致呼吸困难和其他症状。
2. 吸入性糖皮质激素 (Inhaled Corticosteroids, ICS) - 这是一种用于治疗哮喘和COPD的药物类型。它们通过进入呼吸道来减轻炎症和减少过敏反应。

因此，该句话中的两个主要领域术语是“慢性阻塞性肺疾病”和“吸入性糖皮质激素”。

医药领域

TechGPT实例：LLM2KG

问

"汽车发动机是机械工程领域中的重要研究方向之一。发动机作为汽车的核心部件，负责将燃料和空气混合物燃烧转化为机械能驱动车辆。现代汽车发动机包括多个关键组件，如汽缸、活塞、连杆、曲轴等。其中，汽缸是发动机的基本工作单元，用于容纳气体燃烧过程，并将产生的压力转化为线性运动。活塞通过连杆连接到曲轴，将往复直线运动转化为旋转运动。曲轴作为发动机的输出轴，将活塞的运动转化为可用的动力输出。此外，发动机还包括气门、燃油喷射系统、点火系统等辅助部件，以实现燃料的供给、点火和气门的控制。在机械领域的研究中，我们致力于提高发动机的效率、降低排放和噪音，并探索新型材料和技术，以满足不断发展的汽车行业需求。"

机械领域

请抽取出所有三元组。

答

<汽车, 发动机, 主要部分>;<发动机, 类型, 汽车>;<发动机, 是, 机械工程领域中的研究方向之一>;<发动机, 作用, 将燃料和空气混合物燃烧转化为机械能驱动车辆>

TechGPT实例：LLM2KG

问

用烧瓶蒸馏装置模拟固定床脱水设备,研究用苯胺盐酸盐作脱水剂的复盐法将水氯镁石(MgCl2·6H2O)脱水制备无水氯化镁(MgCl2)的过程.考察复盐配料比例、保护气含水浓度、热解温度、热解时间等因素对所制备无水氯化镁产品中氧化镁和水分含量的影响,优化用固定床脱水工艺条件.结果表明,完全脱水和脱苯胺盐酸盐的最佳温度和时间分别是300℃与25～30 min和320～350℃与10～20min.最佳配料摩尔比C6H5NH2·HCl:MgCl2·6H2O为1.1:1.0.在最佳条件下,可得到合格的无水氯化镁产品MgCl298%～98.6%,MgO 0.03%～0.38%,苯胺盐酸盐0.10%～0.12%,水分0.60%～0.81%.

完全脱水的最佳温度和时间分别是多少?

答

最佳的温度是300℃，最佳的时间是25～30分钟。

TechGPT实例：LLM2KG

问 "材料科学是一个研究和开发新材料的学科领域，旨在理解材料的结构、性能和制备方法。材料科学的研究范围涉及从金属到陶瓷、聚合物和复合材料等各种材料类型。在材料科学中，科学家们使用实验室技术和先进的仪器来研究材料的微观和宏观特性。他们通过调控材料的结构和组分，以改变其性能和功能。材料科学的发展对于很多领域都具有重要意义，如电子、能源、医药和环境等。“

请基于上文回答：材料的结构和组分对其有什么影响?

材料领域

答 材料的结构和组成对它的性能和功能有重要影响。例如，不同结构的金属材料具有不同的强度和韧性；有机化合物的分子量和化学键的类型可以决定它们的物理性质（如熔点）和化学反应性；纳米颗粒的形态和大小会影响其在生物医学应用中的生物学效应等等。因此，对于特定应用而言，选择合适的材料和优化其结构和组成是实现最佳性能的关键步骤之一。

TechGPT实例：LLM2KG

问

请生成标题："集成电路（Integrated Circuit，简称IC）是电子信息领域中一种重要的电子器件。它是将多个电子器件（如晶体管、电阻、电容等）集成在单个芯片上的技术。集成电路具有小尺寸、低功耗和高可靠性等优势，广泛应用于计算机、通信、消费电子和工业控制等领域。常见的集成电路包括处理器、内存芯片、功率放大器等。此外，半导体材料也是电子信息领域的核心技术。半导体材料具有介于导体和绝缘体之间的电导特性，如硅和锗等。半导体材料的性能对于集成电路的性能和稳定性具有重要影响。在电子信息领域的研究中，还涉及到电路设计、信号处理、通信协议等多个方面，以推动电子技术的发展与应用。"

答

集成电路与半导体材料——电子信息领域核心技术解析

TechGPT实例：LLM2KG

问

把下面文档里的时间、人物、事件的经过提取出来。

2023年7月30日20时许，房山蓝天救援队接到窦店镇白草洼村一女子求助称，家里积水深约2米，自己带着一个不满3周岁的孩子，特别害怕。救援队立即组织包括王宏春在内的队员前往救援，因途中积水太深无法通行，所有队员回撤。当晚22时左右，王宏春随队配合琉璃河镇政府进行人员安置保障服务。7月31日凌晨3时左右，王宏春再次随队奔赴白草洼村进行人员转运工作，4时多到达班各庄大桥时，发现大桥地基损坏严重，通行存在很大安全隐患，于是王宏春和其他5名队员配合两名路政工作人员留守大桥，拦截劝导人员车辆，另外一组队员徒步绕道3公里山路去往白草洼村转移群众。 31日早晨9时左右，班各庄大桥下的河道突发洪峰，王宏春等6名队员很快被洪水围困，他们和两名路政工作人员一起选择了弃车上船的自救方式，8人迅速穿好救生衣，戴好防护[illegible]，[illegible]逃生。然而，由于山洪迅猛，水流湍急，水中漂浮物太多等原因，导致橡皮艇翻船，船上人员均被急流冲走。接到王宏春等队友落水的消息后，在白草洼村转移群众的6名蓝天救援队员立即赶来搜救。31日上午9时30分，在下游约2公里处，两名落水队员被找到并安全转移到当地医院。11时23分，第三名落水队员被安全救上岸。16时28分，第四名落水队员成功自救上岸。但因通信中断、道路封锁，河岸两边无法全线靠近，搜救队员分多路对所有能靠近的岸边进行排查，同时启用无人机进行空中搜索，均未能找到另外两名落水队员。8月1日15时38分，第五名落水队员王宏春在河道下游约30公里处被发现，被救援上岸，经法医现场确认，已无生命体征。王宏春是房山蓝天救援队第一批队员之一，她热心公益事业，经常自掏腰包贴补队里。2013年刚入队的她就自费4700元购买一套绳索装备，用于队里的山野救援；在此次强降雨抢险救灾中，她安顿好自己6岁的女儿和家人，自费购买水饺方便面等食物送到队里。就在7月31日最后一次出动前，她还提前一小时起床给队友煮饺子…

通用领域

TechGPT实例：LLM2KG

答

时间： 2023年7月30日至8月1日
人物： 王宏春（失踪）、其他队员（幸存者）
事件经过：7月30日晚上，王宏春带领队伍前往营救一名被困在家中的母亲和她的孩子，但由于水深过深，队伍返回了出发点。随后，王宏春和其他队员继续前往营救，但遇到了桥塌的危险，不得不退回原地等待支援。当天晚上，王宏春与其他队员帮助当地政府疏散群众，并在凌晨3点左右再次出发前往营救。但是，当他们在桥上遇到洪水时，橡皮艇翻了，王宏春和其他五名队员被冲走了。搜救队员们进行了多日的搜寻，最终在第11天找到了王宏春的下沉地点，但没有找到另外两个落水队员。

TechGPT实例：KG2LLM

问 请回答这个问题:在房间里有张三、李四、一只猫、一个盒子和一个篮子。张三把猫抱起来放进篮子里，然后他离开房间去上学。当张三不在的时候，李四把猫从篮子里拿出来放进盒子里，然后李四离开房间去工作。最后，张三和李四都回到房间，但他们都不知道在他们不在时房间里发生了什么。这时，他们会怎么想?

答 当张三和李四回到房间后，他们可能会发现以下情况:

1. 猫仍然留在盒子中 如果李四没有将猫放回篮子里，那么它可能还在盒子里。在这种情况下，张三会注意到猫的位置并询问李四是否将它放回篮子里。如果李四承认错误，则他将需要将其重新放入篮子中。
2. 猫被移动了如果李四将猫从篮子里移到另一个地方(例如桌子上)，那么张三会发现猫的位置不对，并开始寻找答案。他会问李四是否将猫放在某个特定的位置或做了什么其他事情。
3. 没有猫-如果李四将猫从篮子里拿走了，而张三不知道这一点，那么当他找到空荡荡的篮子时，他很可能会感到困惑。他可能会开始搜索整个房间以确定猫的行踪。如果他最终无法找到猫，他可能需要向其他人寻求帮助或者打电话给警察报告失踪的动物。

文本语义理解、逻辑推理能力强!

TechGPT实例：KG2LLM

问 写一个“美丽肤”熬夜面膜的营销广告。

答

标题：拯救你的肌肤，尝试我们的美丽肤熬夜面膜！

正文：

每天都要应付工作、学业和生活中的种种压力，加上现代人的不良生活方式，导致让肌肤变得越来越无法自愈。而在这样的情况下，熬夜更是对肌肤造成了极大的伤害。因此，为了保持健康的肌肤，我们特制定了一款名为“美丽肤”熬夜面膜的产品，希望能给你提供最好的解决方案。

该面膜主要由多种天然有机成分组成，包括绿茶、芦荟、海藻等，可以深入清洁肌肤，去除污垢和油脂，同时还能修复肌肤，使其更加清新、亮丽。此外，该面膜还添加了多种保湿成分，如玫瑰花水、胶原蛋白等，可以长期保持肌肤的水分平衡，从而达到淡化皱纹、改善肤色等效果。

使用方法也非常简单，只需将面膜敷于脸部15-20分钟后取下即可。建议每周使用3次，并配合日常护理程序进行使用，相信你会发现肌肤状态的显著改善。

最重要的是，我们的面膜完全安全无副作用，没有任何化学物质或香料，所以你可以放心地使用。让我们一起选择“美丽肤”熬夜面膜，让你的肌肤再度焕发青春光彩！

文本生成逻辑严密！

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

知识图谱研究的发展趋势

❖趋势1：知识图谱**基础研究任务**将更加活跃！

- “**大模型+行业**”是可行的大模型落地模式
- 各类**行业数据**需精细化分析与加工➔形成**行业知识图谱**
- 行业标注数据少➔**低资源场景**下知识图谱构建将成为重要研究内容
- ……

知识图谱研究的发展趋势

❖趋势2：知识图谱的内涵将不断扩展！

- **多元关系**知识图谱构建
- **多模态**知识图谱构建
- **事理**图谱构建
- **低资源场景**下的**行业**知识图谱构建
- ……

知识图谱基础研究任务将相应扩展！

知识图谱研究的发展趋势

❖趋势3：大模型与知识图谱相互促进

- 知识图谱帮助大模型成为**大知识模型**或**大智慧模型**
  - 结果可**解释**、可**溯源**
  - **可靠**结果生成
  - 知道自身**能力边界**
  - ……

知识图谱研究的发展趋势

❖趋势3：大模型与知识图谱**相互促进**

- 大模型帮助知识图谱提升**基础研究任务**的效率与性能
  - 知识图谱构成要素抽取的**性能**更强
  - 知识图谱要素抽取的**泛化能力**更强
  - Scheme构建**自动化**
  - ……

主要内容

- 知识图谱发展简单回顾
- 大模型对知识图谱研究的挑战
- 知识图谱研究中的传承与变革
- 知识图谱研究的发展趋势

结束

谢谢大家！