



勘探开发知识图谱计算技术 研究及应用

谭培波
2018年8月



谭培波

——北明智通知识工程事业部技术总监

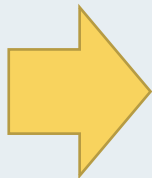
高级工程师，国际TRIZ认证三级专家，6sigma黑带大师，并担任国家创新方法研究会企业分会理事等职务。

谭培波先生具有10年卫星通讯系统研发经验，参与多个卫星型号研发，获得3项部级科技进步奖；10年移动通讯系统研发经验，并推进中兴集团6sigma工作，申请3项基站BTS系统和多块单板开发相关突破性改进专利。十余年来带领技术团队开展知识图谱构建与应用研究，铸就北明智通语义魔方与知识工程平台两大旗舰产品技术内核，并申请3项语义分析相关技术发明专利。

人工智能标准化白皮书 (2018 版)

指导单位：国家标准化管理委员会工业二部

编写单位：中国电子技术标准化研究院
二零一八年一月



3.1.2 知识图谱

知识图谱本质上是结构化的语义知识库，是一种由节点和边组成的图数据结构，以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体—关系—实体”三元组，以及实体及其相关“属性—值”对。不同实体之间通过关系相互联结，构成网状的知识结构。在知识图谱中，每个节点表示现实世界的“实体”，每条边为实体与实体之间的“关系”。通俗地讲，知识图谱就是把所

13

有不同种类的信息连接在一起而得到的一个关系网络，提供了从“关系”的角度去分析问题的能力。

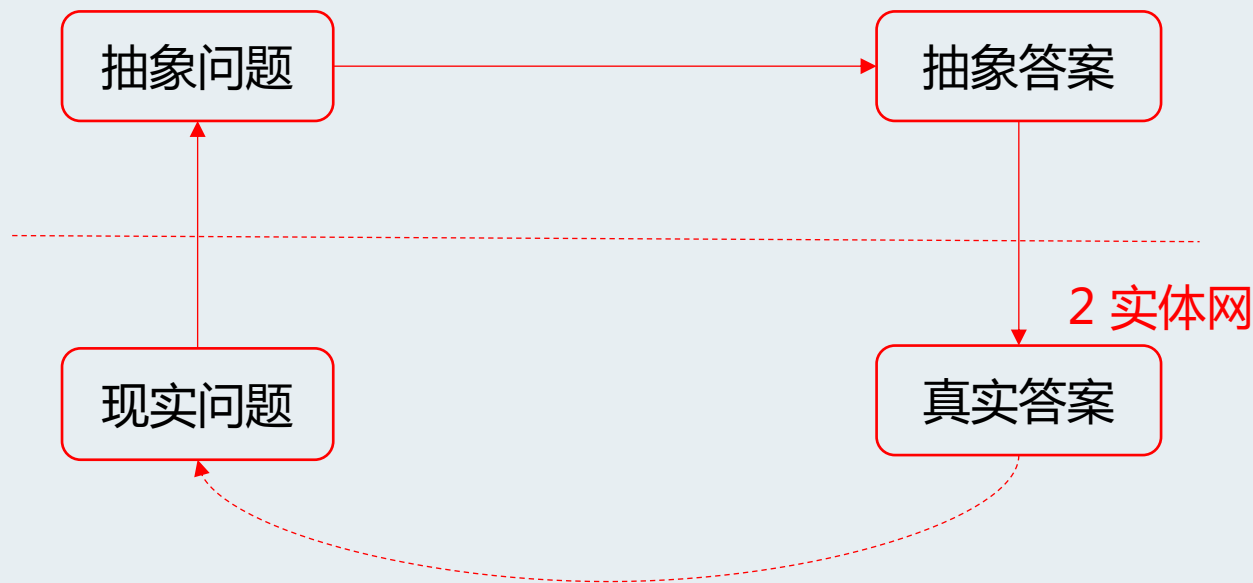
1.2 我们的定义

代数
化学反应
技术矛盾
1 语义网、概念图谱

解方程
生成物表达式
创新原理
最短路径、生成树



抽象、大脑、
概念、意识



现实、实体、
物质性

影响概念网的因素
1 概念网上点的规模，5000个概念还是100个概念，也就是业务的粒度；
2 概念的抽象程度，家谱的概念网，父子关系还是始祖往下，自关系和他关系；
3 实际中的应用情况，比如日期，虽然是4种标准实体，但是不一定需要到这个粒度。

1.3 知识的图+谱

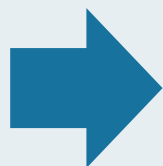
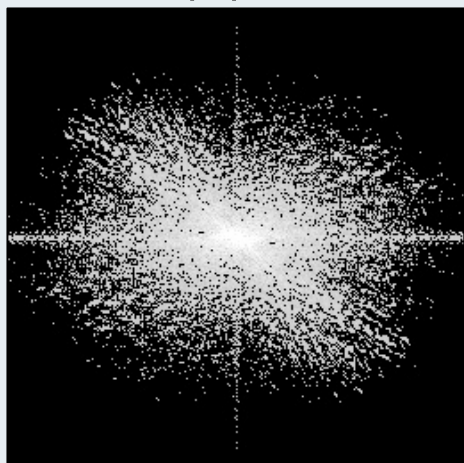
文本处理

知识表达

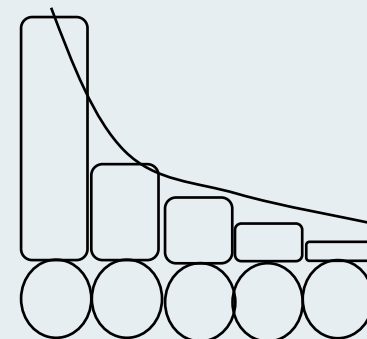
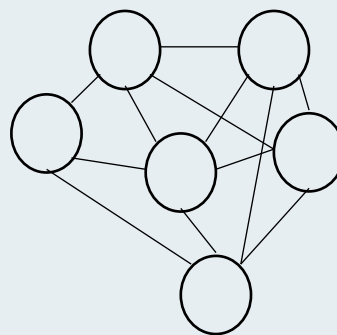
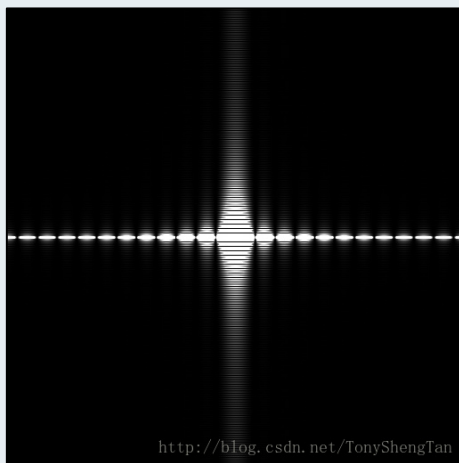
知识挖掘

知识图谱

图



谱



- 再看看光粒子网和光场谱；
- 波粒2相性，粒子性和分布性；
- 生物学家讲量子纠缠，语言学家得讲点基因突变

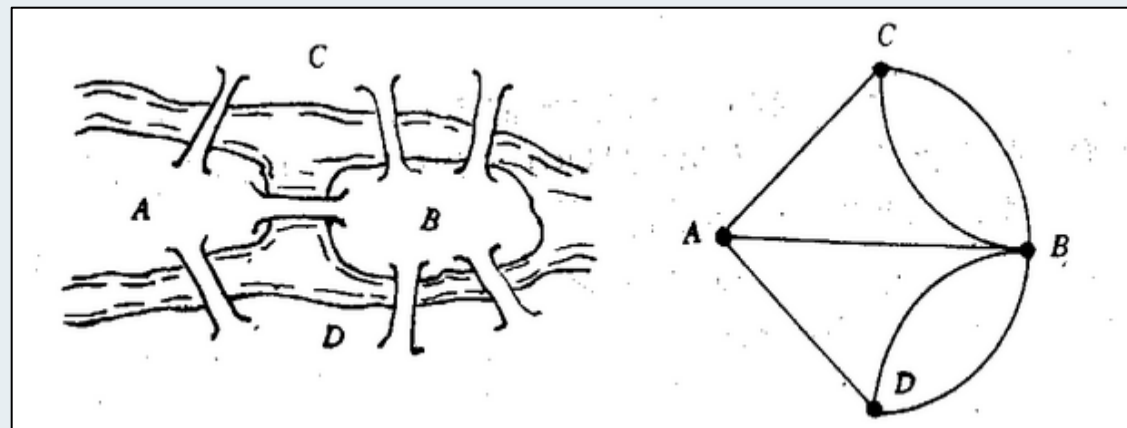
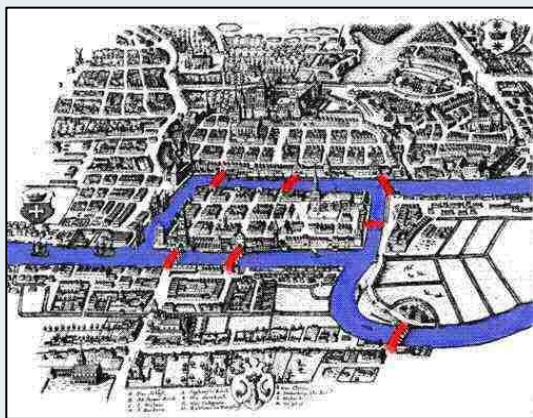
2012年5月，谷歌正式发表自己的知识图谱



2 知识图谱的图=图论

图论〔Graph Theory〕以**图**为研究对象，图是由若干给定的点及连接两点的线所构成的图形，描述事物之间的某种特定关系，点代表事物，线表示相应两个事物间具有这种关系。

拟阵理论、**超图理论**、**极图理论**，以及代数图论、拓扑图论



- 图和网络的基本概念：无向图、有向图、完全图、二分图、子图；
- 直径、平均距离、簇系数与度分布
- 点线面算法，社会网络

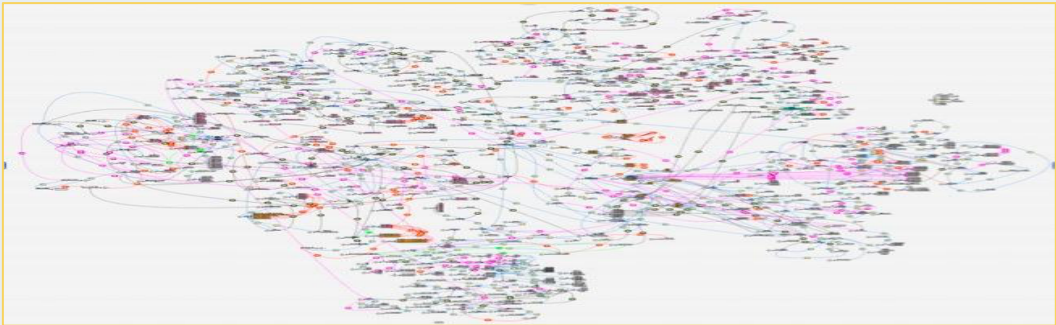
本体论 (Ontology) 存在

实例知识图谱

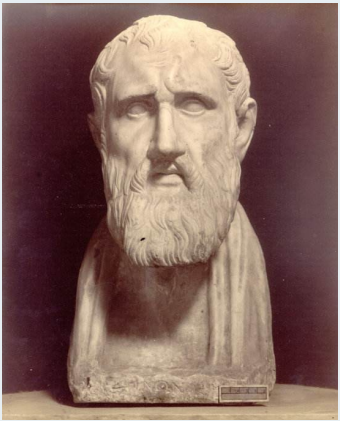
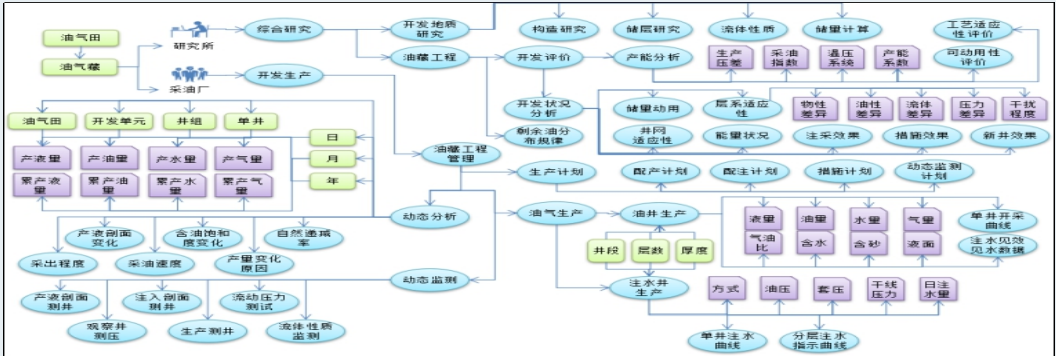
概念语义模型

- 工程中概念和实体的假设，操作层面解决红不红动不动的问题；
- 本体论的理论依据是集合论，结论有哥德尔不完备性定理的悖论；

油气田开发生产知识图谱示例



语义识别模型

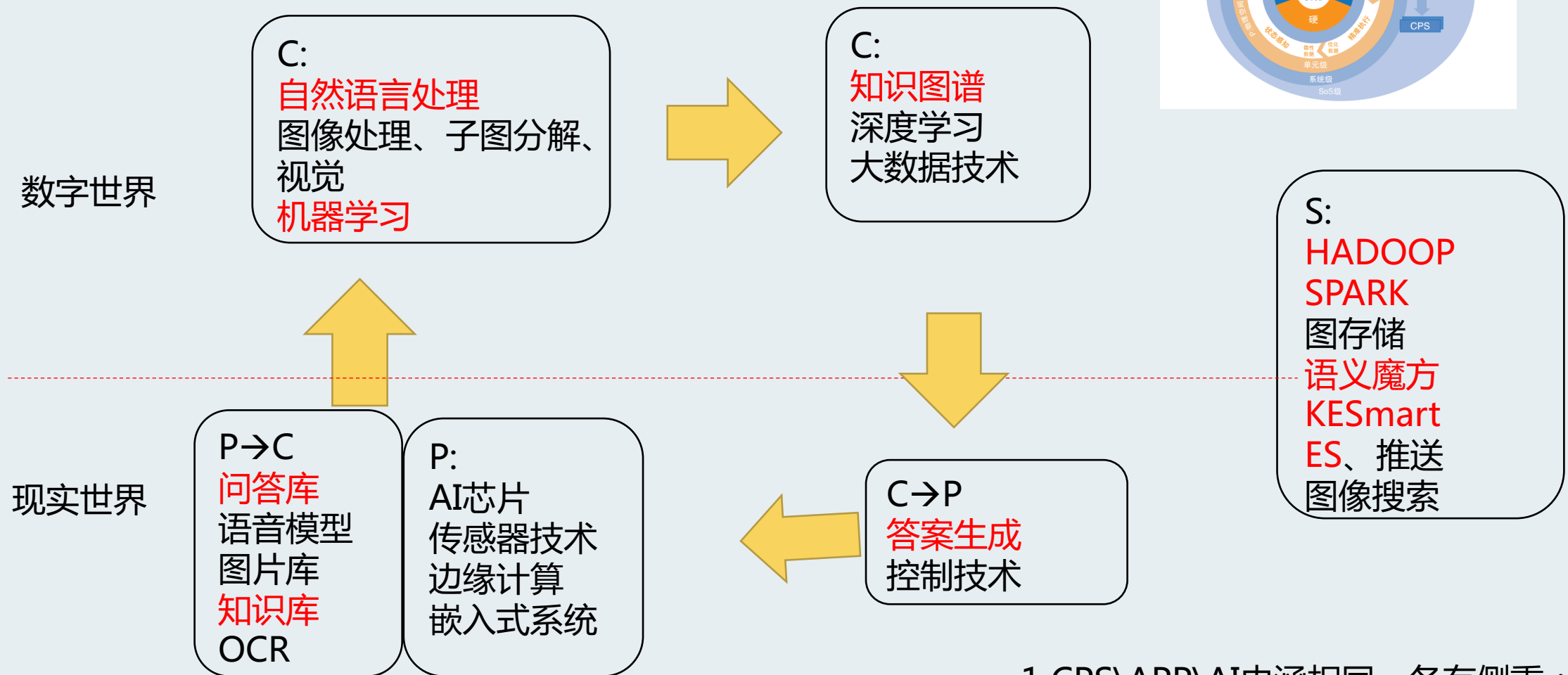


巴门尼德



哥德尔

4.1 基于知识图谱应用逻辑



- 1 CPS\APP\AI内涵相同，各有侧重；
- 2 工业云和AI解决实时问题

■ 认知的过程



■ 图谱解决不了认知精确性的问题

塔里木盆地是大型封闭性山间盆地，地质构造上是周围被许多深大断裂所限制的稳定地块，盆地地势西高东低，微向北倾

认知—是人对事物的各个属性、各个部分及其相互关系的综合的整体的反映。

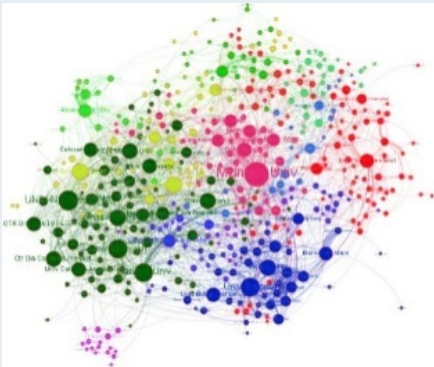
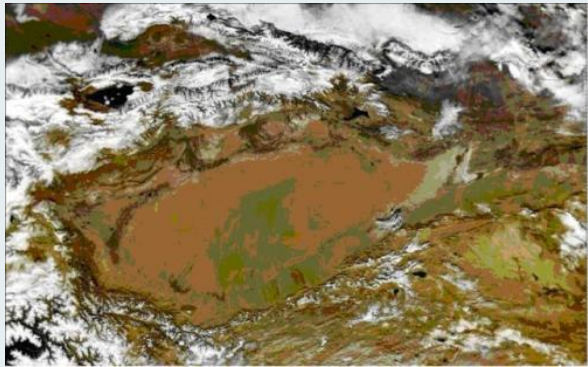
共同特征:

- 认知的整体性
- 认知的理解性
- 认知的恒常性
- 认知的选择性

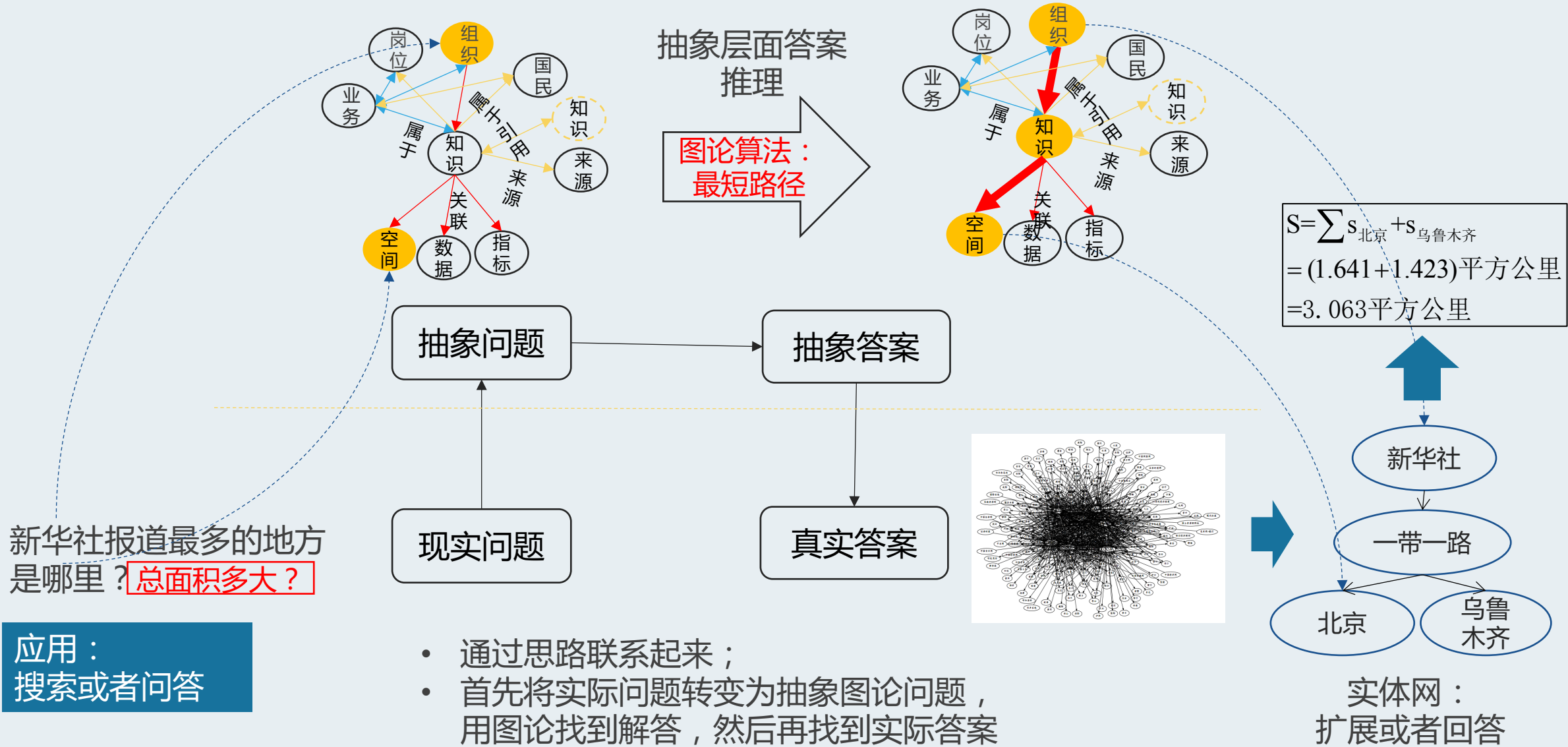
- 认知的科学性
 - 认知的精确性

■ 图数是一种语言，数和其它语言结合，就能解决认知精确性的问题

塔里木盆地南北最宽处**520**千米，东西最长处**1400**千米，面积约**40**多万平方，海拔高度在**800-1300**米之间。
由于数字的内容可以与其他形式进行表达且不容易出错，如次数、频率、幅度，所以数字也是全球最易理解的文字或符号。



5 智能问答举例-精确性



6 基于图谱计算的路径



概念与实体的相对性，粒度的适应性

输入转换FAQ

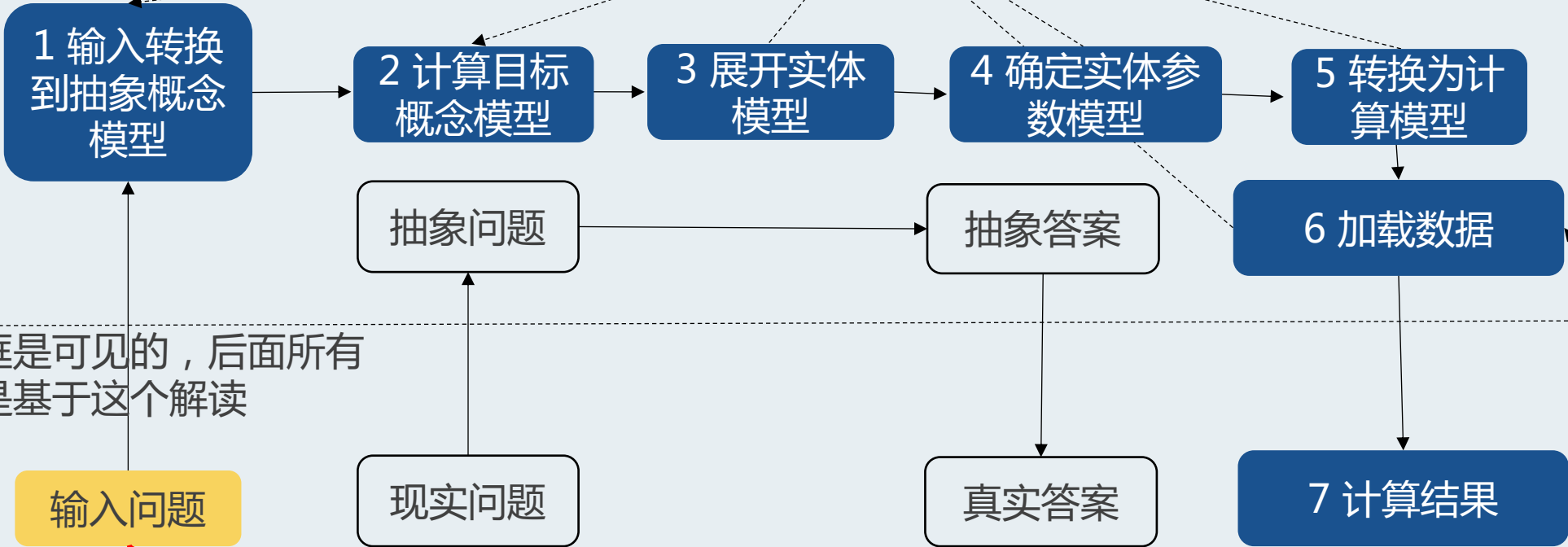
(1) Q转换

(2) 概念目标模板

(3) 实体模型模板

(4) 参数模板

(5) 数据加载模板



只有输入框是可见的，后面所有的计算都是基于这个解读

6.1 问题分解模板



问句	答句	问句类型 1：查表 2：统计	问句意图分析		业务和对象		业务或者对象的属性	
			框架名	taget	A：业务和对象分类	B：业务对象特征词	属性名	X：属性特征词
X井井位在哪里		1	钻井工程	井位	井	井位	井位坐标	在哪里
X井井位坐标多少		1	钻井工程	井位坐标	井	井位坐标	井位坐标	多少
X井是什么井		1	钻井工程	井是什么	井	什么井	井别	什么井
X井是干什么用的		1	钻井工程	井是干什么	井	什么用	井别	什么用
X井是干什么的		1	钻井工程	井是干什么	井	干什么	井别	干什么
X井的井别是什么		1	钻井工程	井的井别	井	井别	井别	是什么
X段井深多少		1	属性	段井深	层	井深	井深	多少
X盆地有多少凹陷？		2	统计分析	有多少	盆地	盆地	凹陷	凹陷
新疆有几个盆地？		2	统计分析	有几个	地区	新疆	盆地	盆地
X井区块有几口井？		2	统计分析	有几口井	区块	有几口井？	井	有几口井？

6.1 问句解析示例



问句：孤岛油田含油面积是多少？

匹配上的模板：.*油田含油面积是多少

问句模板

查询语句=MATCH p =(x:OilField { 名称:"孤岛油田" })-[r:包含]-(y:OilUnit { }) RETURN sum(toFloat(y.含油面积)) as key

回答句模板=孤岛油田含油面积是Query_Result平方千米

答句模板

问句：孤岛油田1月产液量大于100方、含水大于95%的井有多少？

匹配上的模板：.*油田.*月产液量大于.*方、含水大于.*的井有多少

查询语句=MATCH p =(x:OilField { 名称:"孤岛油田" })-[:包含]-(:OilUnit)-[:包含]-(y:OilSpring)-[r:OilMonth]-(z:OilTime {名称:"201801"})

where (toFloat(r.月产油量)+toFloat(r.月产水量))>100

and (toFloat(r.月产水量)>0 or toFloat(r.月产油量)>0)

and (toFloat(r.月产水量)/(toFloat(r.月产油量)+toFloat(r.月产水量)))>0.95

return count(y) as key

回答句模板=孤岛油田1月产液量大于100方、含水大于95%的井有Query_Result口

答句模板

最短路径法

对于两个点的解析，采用最短路径法。

最小生成树

对于多个点的解析，采用最小生成树。

所以要采用树的结构，
是因为**层次结构易于转化为计算模型**

基于图谱的问答的基本出发点，是把问题等效为对图的结构和图的属性的认识

具体分为：

1、关于点的问题，等效为对于点的定义和属性的认识，如

- 川22井有几个属性？这是求A的出度
- 川22井属于几个油田？这是求A的入度

2、关于线的问题；最短路径

2.1 第1种情况：A-ab-B存在

关于线的构成和结构的问题在2点一线中，知道2个要素求第3个要素的值，如

- 川2井的位置在哪里？这是已知A-ab求B=？
- 谁的位置在四川？这是已知ab-B求A
- 川22井在四川吗？这是已知A-B求ab

2.2 A-ab-B之间不存在关联，等效为最短路径
川22井在云南吗？这是A-B求Xab

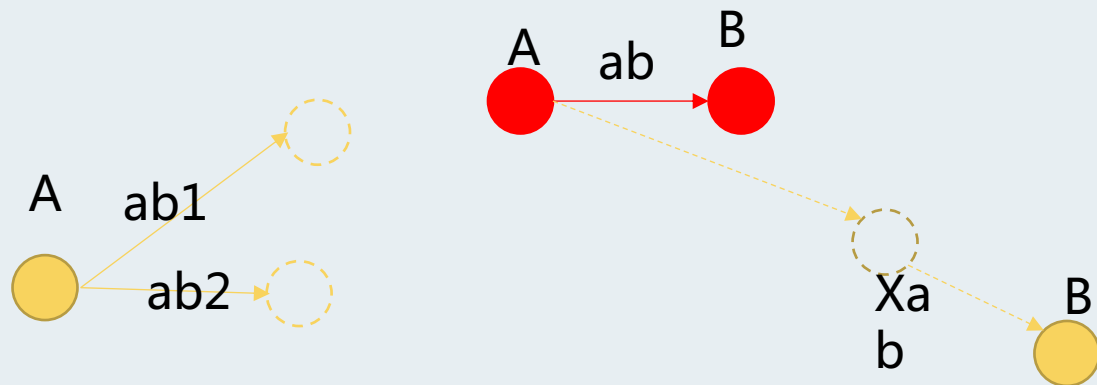
2.3 关键点

塔里木盆地和柴达木盆地关联的最重要的属性是什么

3、关于面的问题

3.1 最小生成树

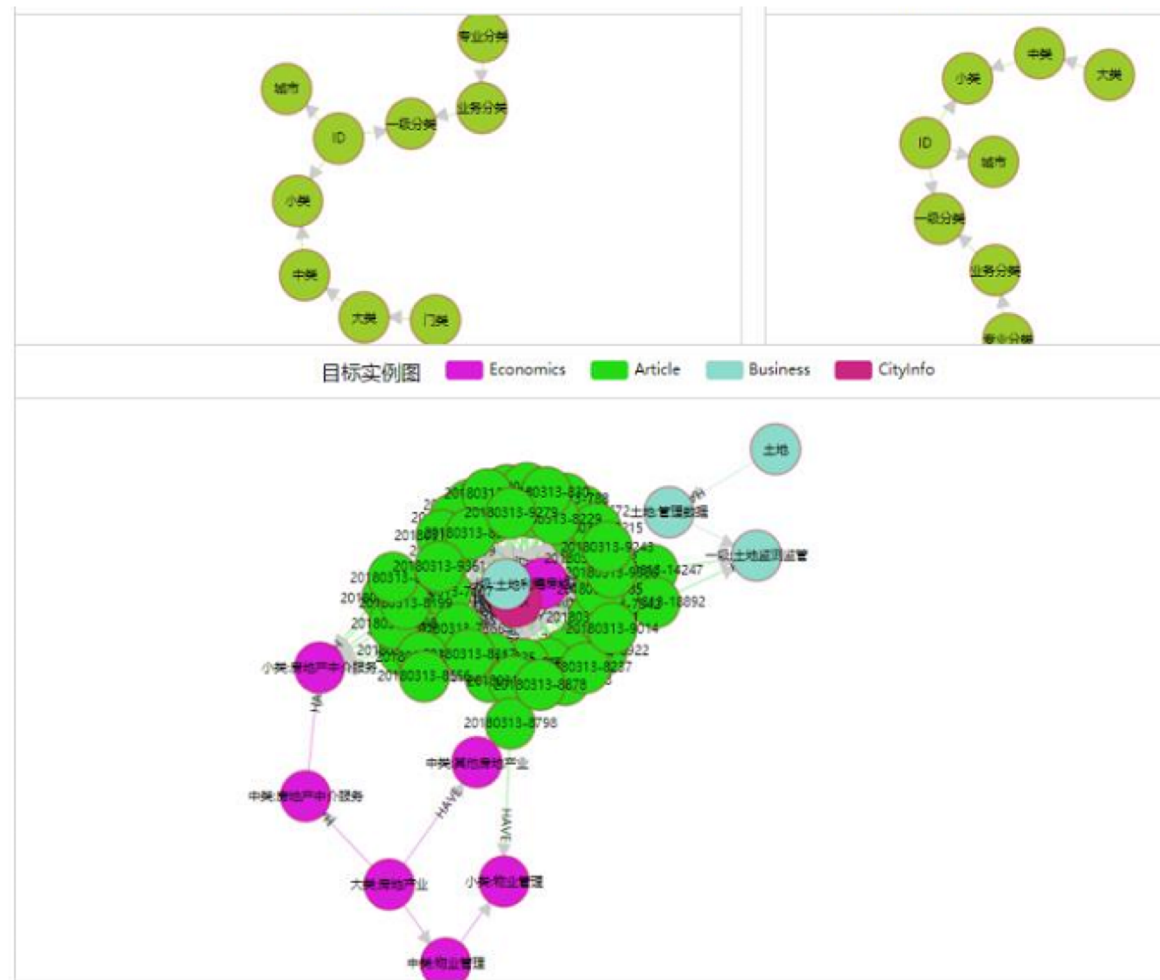
- 川22井的知识图谱的核心骨架是什么
- 在北京讲农业的新闻也在讲房地产啊



将按照模板

将概念网转换为实体网；

一个例子如右图所示，图上方是概念网的情况，下方是实体网



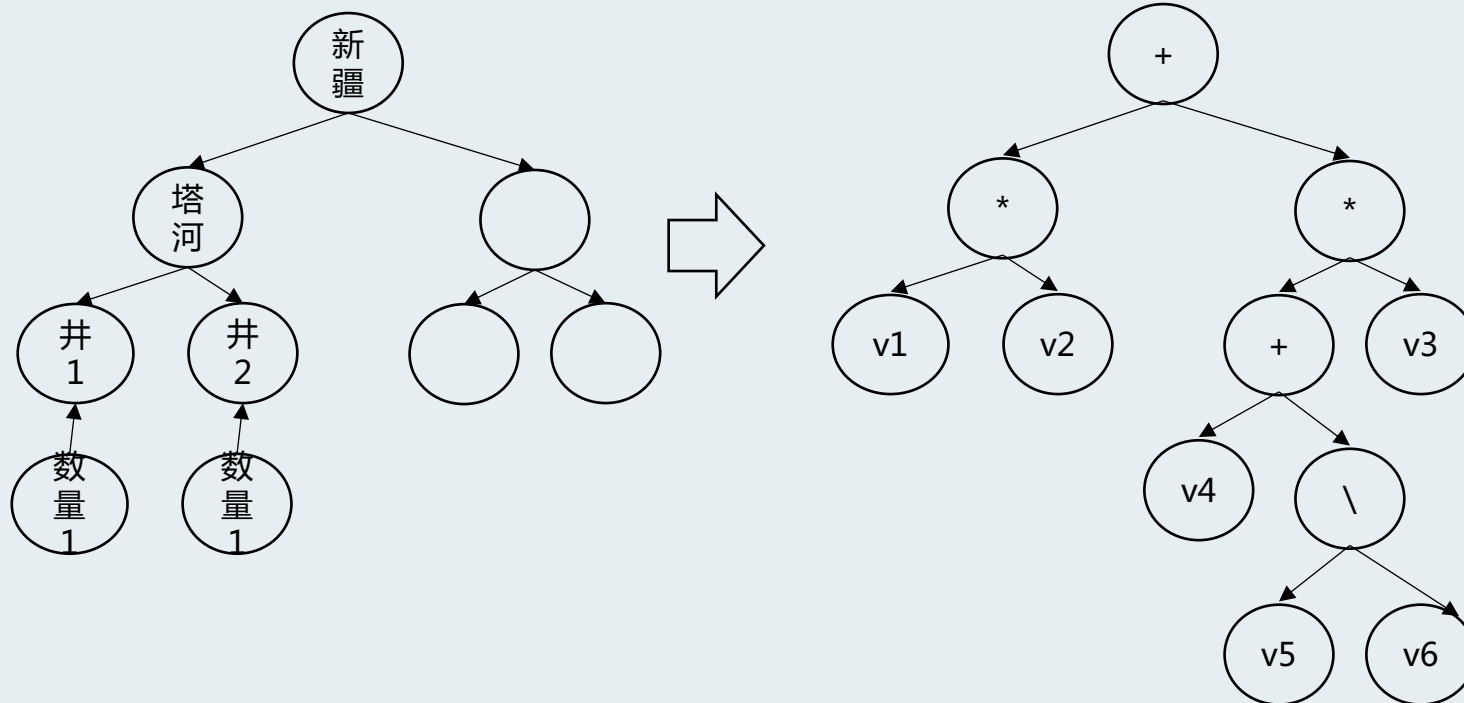
6.4 将参数模型转换为树计算模型

统计方法

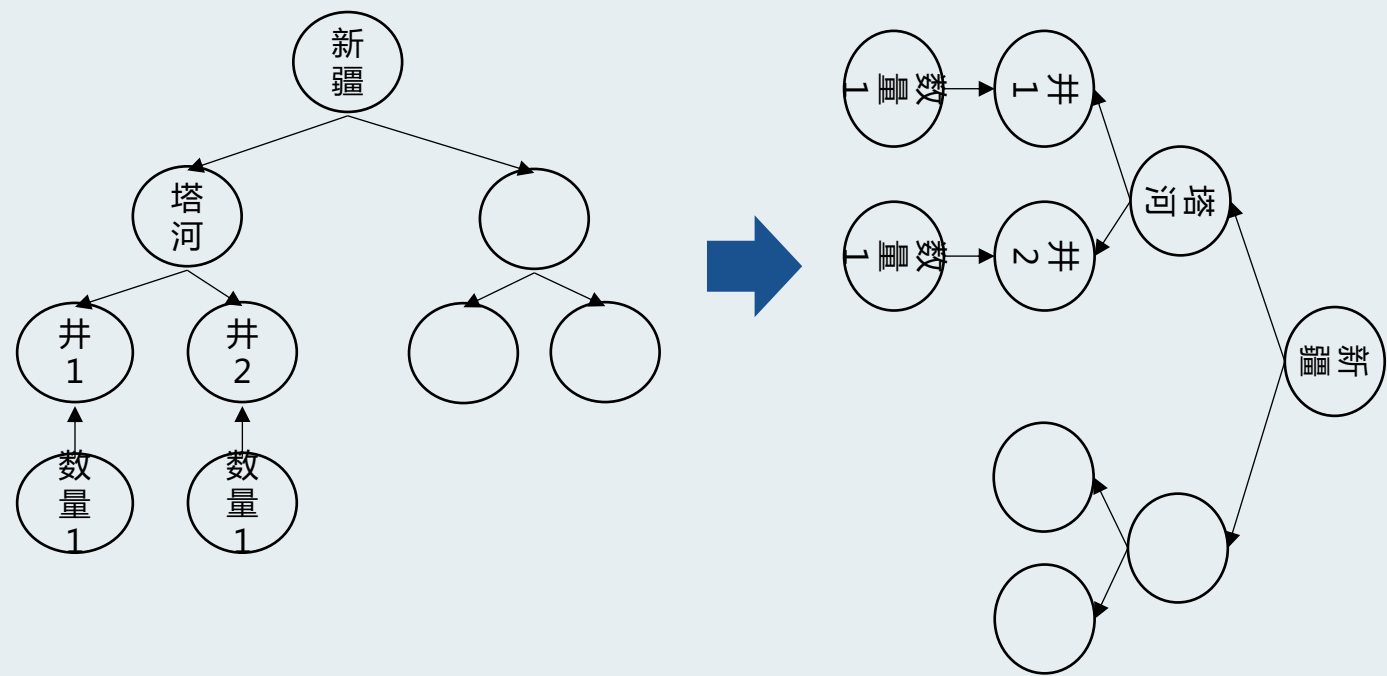
计算树方法

注意

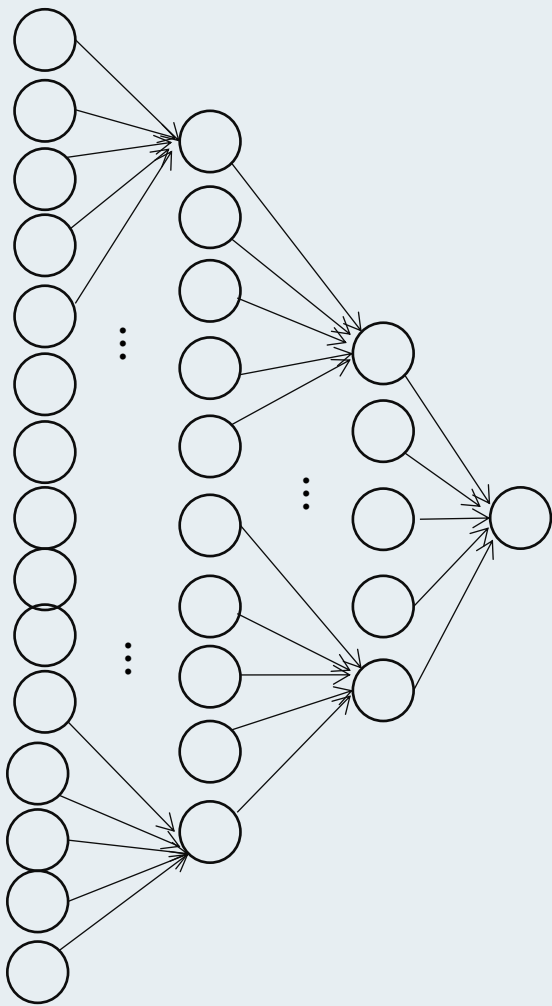
- 1 在计算的初级阶段，以树为基础，对于网的计算，还需要更多的场景探索；
- 2 网意味着不同路径到达同一个点，隐含着一致性检验问题



深度学习方法



深度学习模型可以解决一些复杂的训练问题



7 计算结果举例

精准问答：从油田海量生产信息中直接给出想要的答案，推理出问题原因和解决方案，提高巡检人员问题发现与解决效率。



知识图谱在不同
行业应用的扩展

语义的图论解析
方法研究

GAN在知识图
谱构建中的应用

谢谢！



关注我们，请扫二维码

北明智通（北京）科技有限公司

地址：北京市西城区德胜门外大街13号院1号楼 合生财富广场2层

电话：(86) 10-57965669

传真：(86) 10-57965677

邮箱：info@bmsmart.com.cn

官网：<http://www.bmsmart.com.cn>

THANKS



Know 2 Smart