

数智科技创新推进司法现代化

潘 理

信息内容分析技术国家工程研究中心
上海交通大学

2024年11月23日

上海 华东政法大学



目录



01 背景意义

02 难点挑战

03 路径方法

04 应用案例

数智科技辅助司法是发展趋势

加拿大：iCase系统

办案流程节点管理、案件监督与风险评估等

荷兰：数字化案件监督管理系统

案件与办案人员监督、办案流程节点管理等

日本：在线庭审系统

庭审流程管理、案件与办案人员监督

美国：CM/ECF系统; COMPASS系统

案件监督与风险预警、人员监督等

欧盟：ERP案件管理系统

事件监督与风险评估、办案流程节点管理等

新加坡：ICMS系统

办案人员监督预警、案件风险评估等

我国司法体系在规模和复杂性方面居世界前列

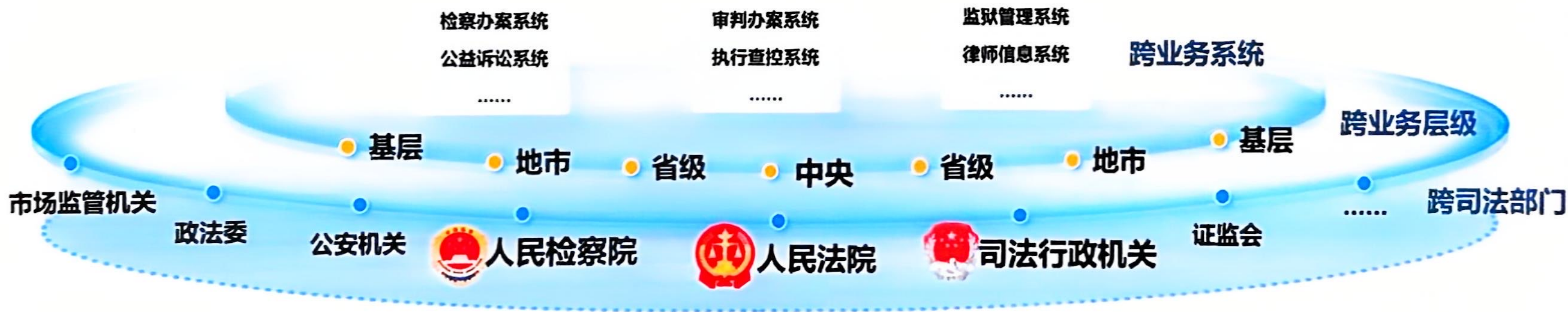
数据

海量多元数据
数千家机构

系统

数千万用户
数千种系统

业务

亿级案件规模
数千业务场景

数智科技创新推动我国司法现代化建设

司法领域知识引导的智能应用

多元业务数据深度融合与挖掘

业务需求

技术发展



全面贯彻落实公正司法

检察业务

审判业务

司法业务

大数据
分析人工
智能

区块链

2020

2024

2025

2035

已有基础

司法信息化3.0

建设现状

智慧司法深入发展

“十四五”规划

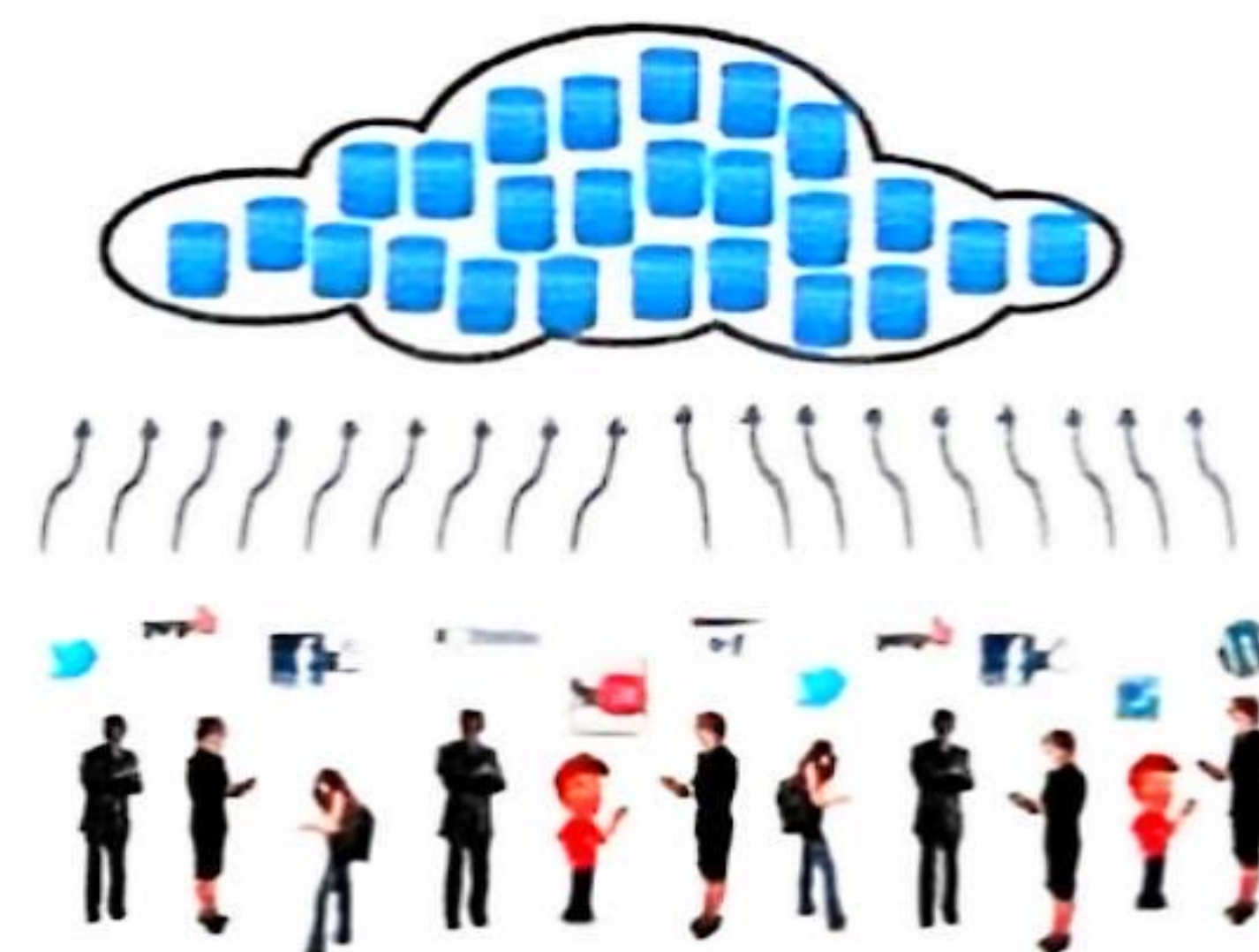
全面实现智慧司法

实现司法现代化
法治社会

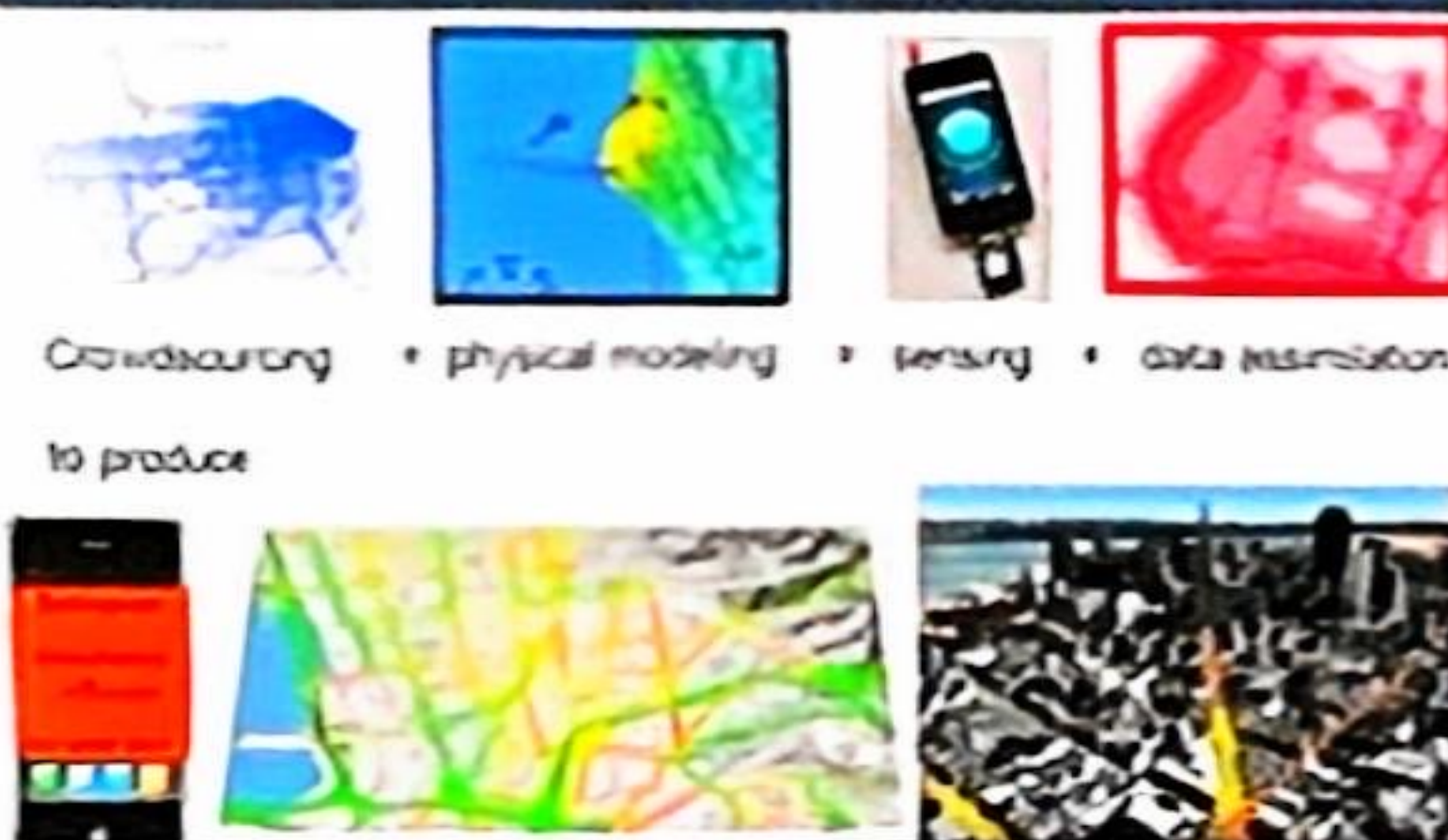
深入开展数智司法科技创新的挑战

- 跨部门多元数据融合不充分、不全面
- 司法业务模型构建困难，可解释性差
- 利用司法领域知识引导增强不足

User Generated Content



NextGenMaps



Graph Data



Internet of things

M2M = Machine to Machine (Telemetry)



数据：共享效能与隐私保护平衡难

- 司法领域人工智能系统涉及不同敏感层级的隐私信息
- 面向的用户使用意图多样

系统：智能算法透明公正决策难

- 人工智能系统算法普遍可解释性差，导致“算法黑箱”问题
- 决策算法会学习并放大司法训练数据中的偏见，导致系统性不公平

业务：合规与全流程质效优化难

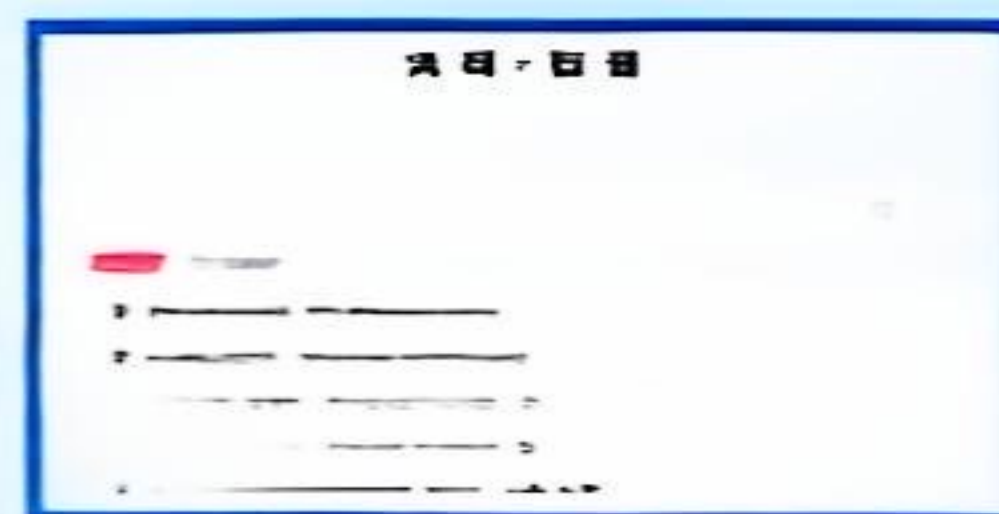
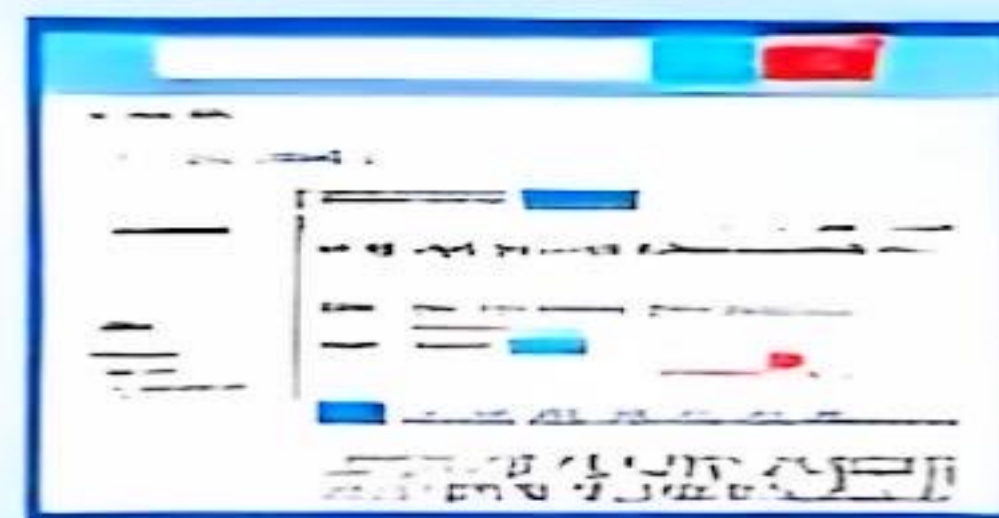
- 司法领域人工智能系统难以确保生成内容合法合规
- 面向复杂上下文语义场景的推理推断

应用：安全态势复杂与监管预警难

- 缺乏针对司法领域人工智能系统独有安全风险的监管技术手段
- 安全问题难以及时发现和应对

司法领域的人工智能系统**数字法院：**

- 法律智能交互问答系统
- 相似案例智能推送系统
- 诉讼风险智能评估系统
- 案件风险智能识别系统
- 规范量刑智能辅助系统

**智慧检务：**

- 检察办案智能辅助系统
- 公益诉讼智能辅助系统
- 检察知识服务类案系统
- 案件质量评查智能系统

**智慧司法行政：**

- 智能法律服务检索系统
- 社区矫正智能服务系统



数字法院典型场景：全流程审判智能化监督

e租宝案

涉案金额大、人员多、关系复杂、执行难

**重要办案节点预警
监督预判不足**

多模态数据预警难、审判执行周期预判难

云南孙小果案

罪行恶劣、保护伞多、多次违规减刑

**重大违规办案事件识别
预警不足**

多通道信息识别难、事理研判预警难

南京彭宇案

案例典型、社会关注度高、舆论热点

**重点案件办案风险监管
预警不足**

多维度风险评查难、全流程案件监管难

海南百亿法官案

受贿、违规经商、违规干预案件

**人员廉政风险
关联预判不足**

廉政风险线索发现难、智能关联预警难

数字检察典型场景：法律监督线索挖掘

石化虚开百亿增值税
专用发票案新型虚开手段不断更新
、监管困难跨数据权属线索
可信传递融合难嵊州市人民检察院
“非标油”治理多部门协作挖掘线索
操作方式复杂跨域跨部门推广
模型泛化执行难绍兴市医保基金
诈骗类案监督个案总结到类案监督
数字监管手段待增强领域法律监督
规则智能提取难华中科技大学同济
医院骗保案医疗行为专业、依赖行
业人士检举风险行为复杂
主动线索生成难

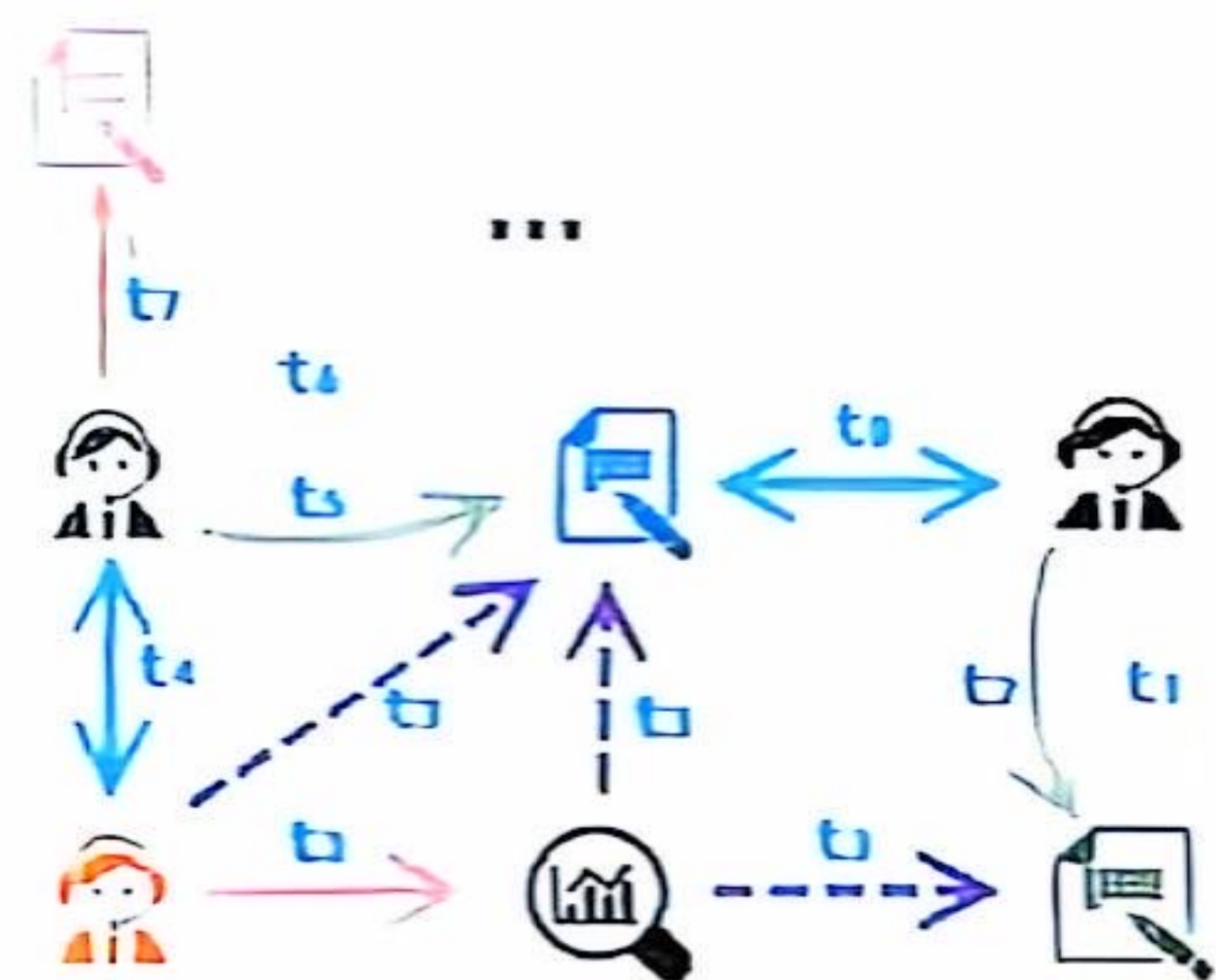
数智司法应用技术发展——从量化到智能



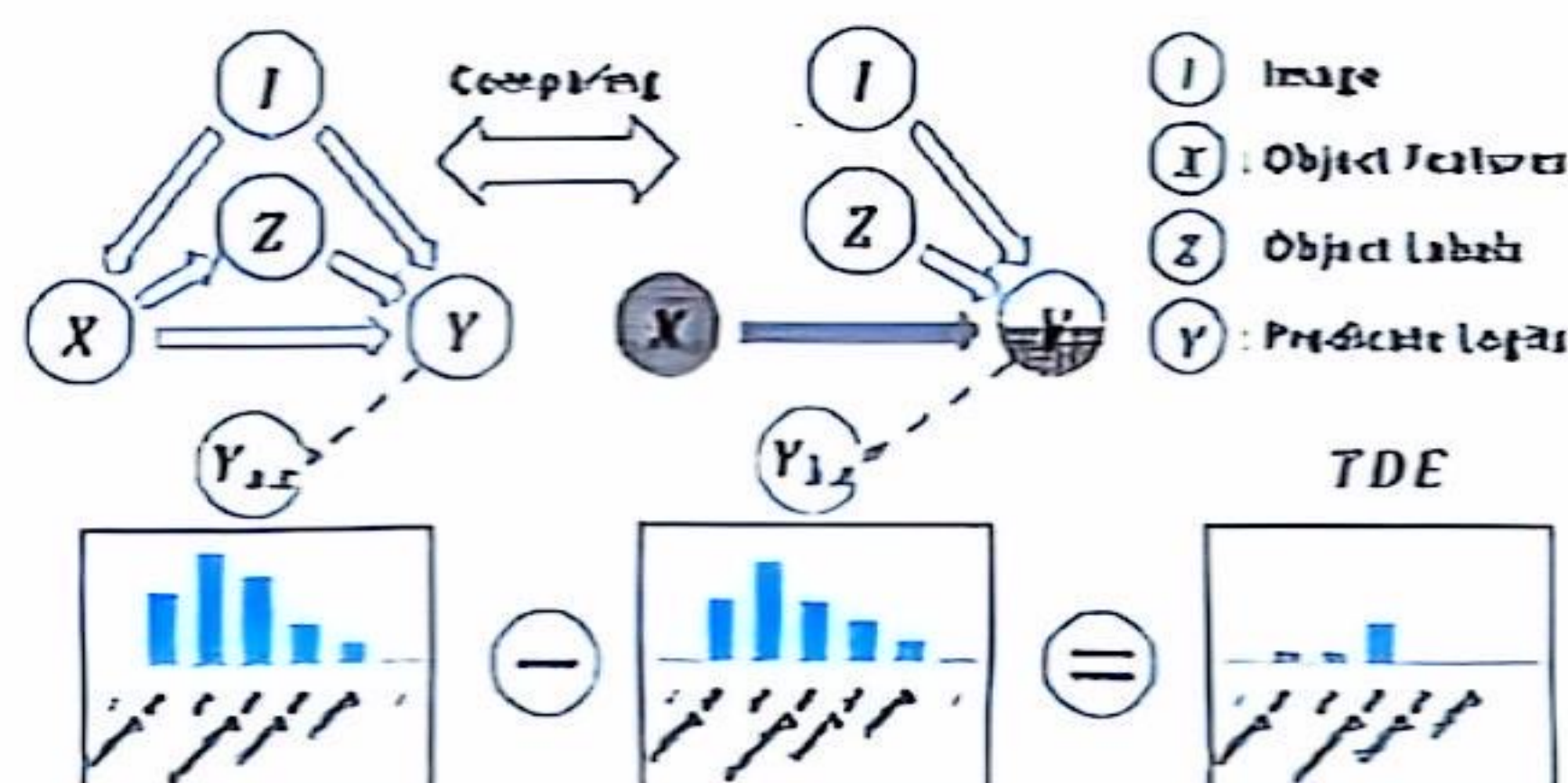
发展可信可用的数智司法创新技术

- **五项基本原则：安全合法，辅助办案，公平公正，透明可信，公序良俗**
- **司法数智技术创新目标：司法认知计算**

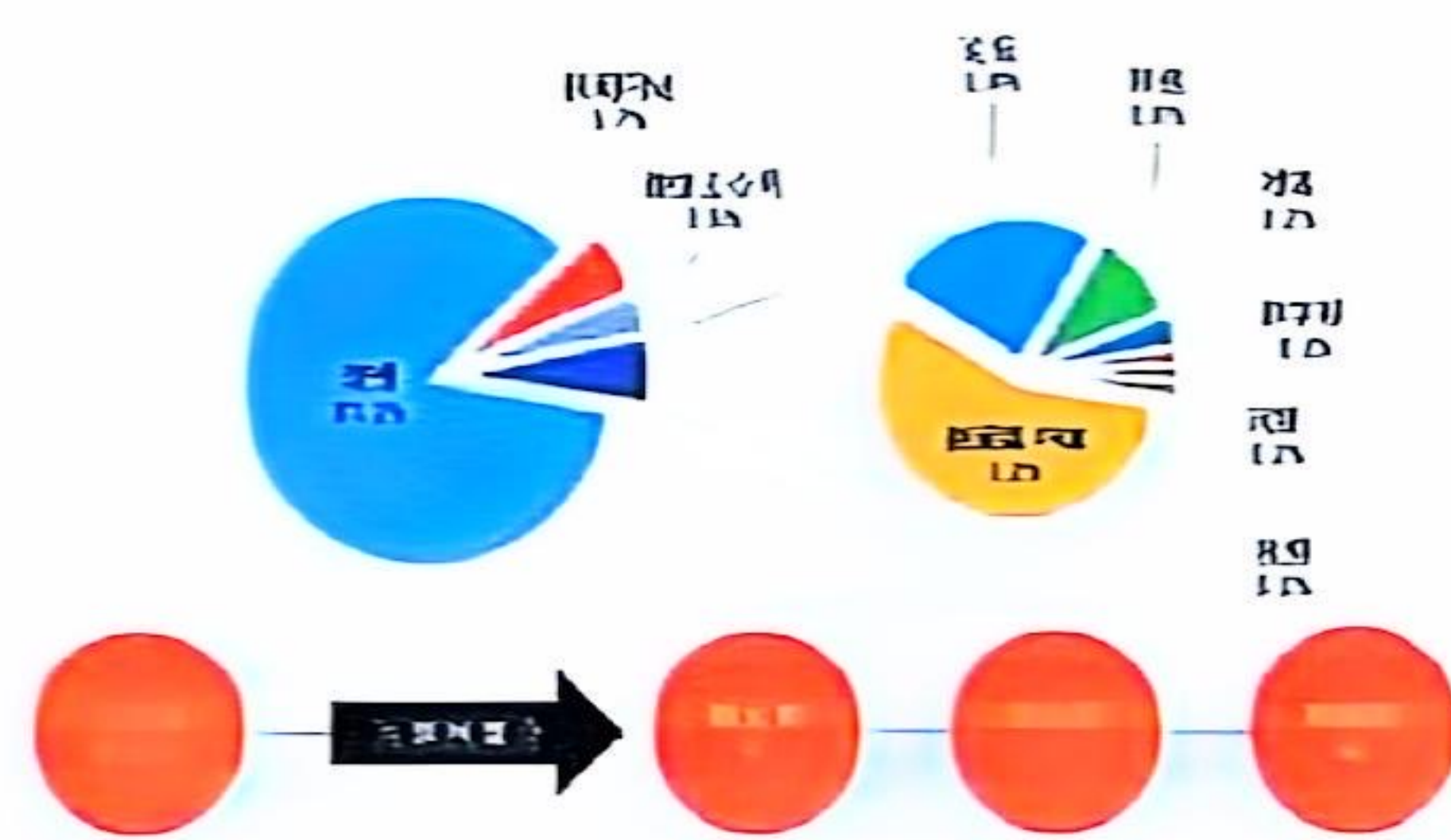
司法大数据分析与**人工智能有机结合**，从非结构化大数据中感知理解司法复杂系统的内在规律，遵循司法领域知识引导，辅助与优化司法业务办理。



行为建模



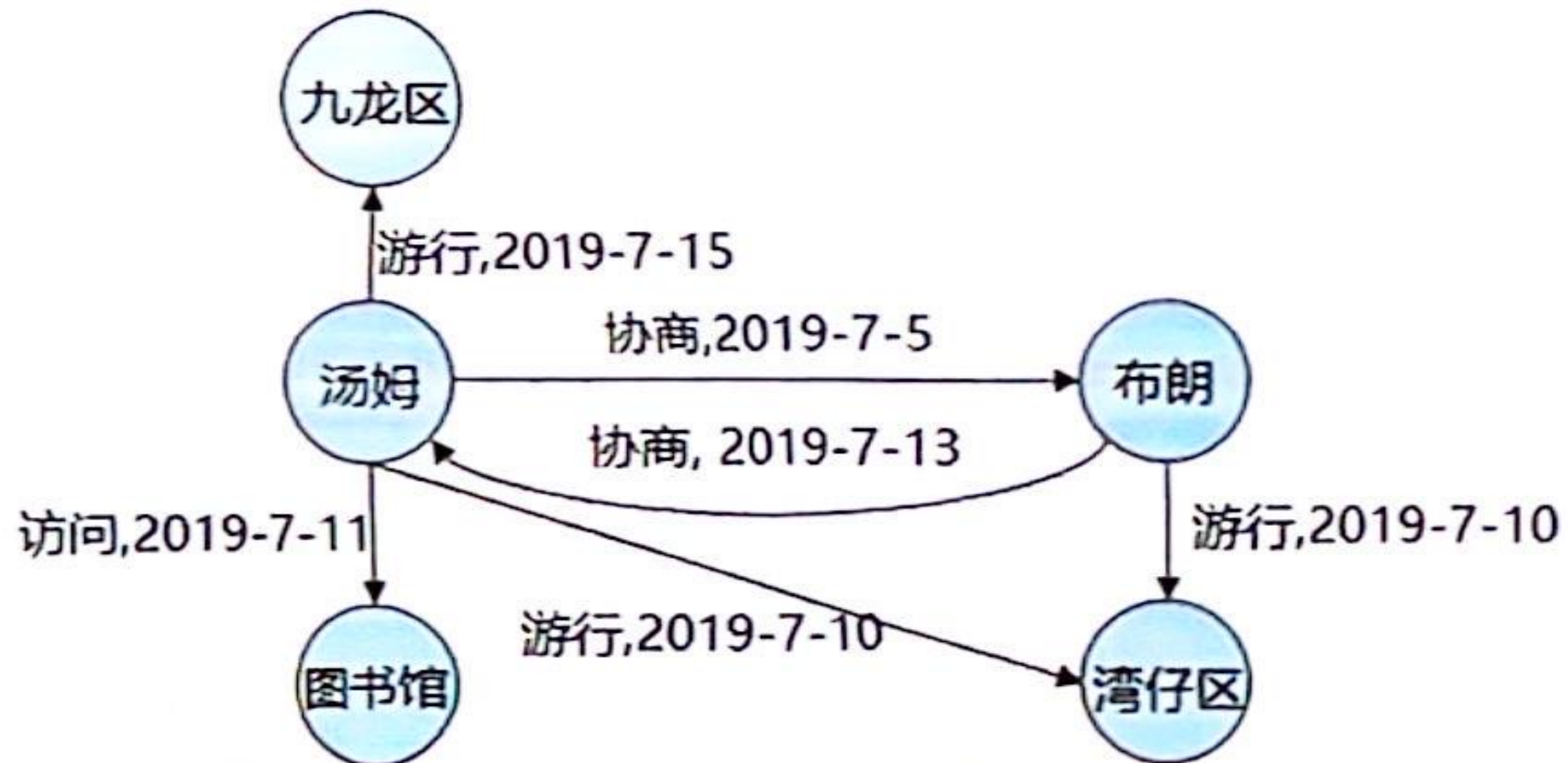
行为因果图构建



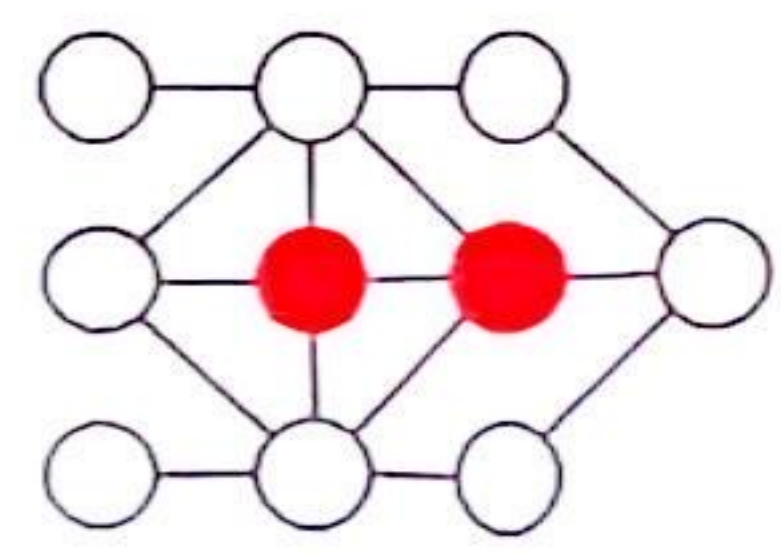
事件逻辑链条提取

● 时序知识图谱及其表征

时序知识图谱是一种扩展了**时间维度**的知识图谱。它不仅包含实体及其关系的信息，还包括这些关系在时间上的变化。



● 时序知识图谱推理应用场景



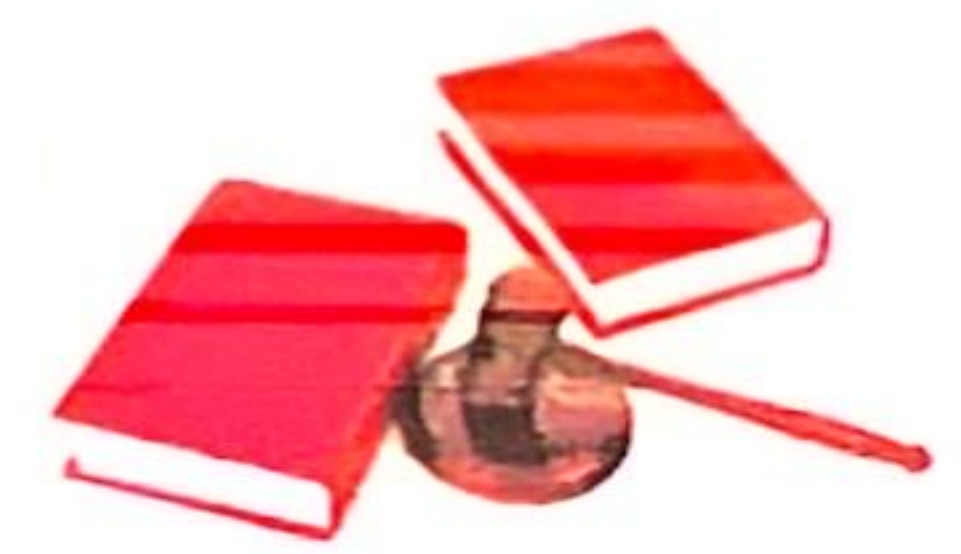
用户行为预测



司法知识问答



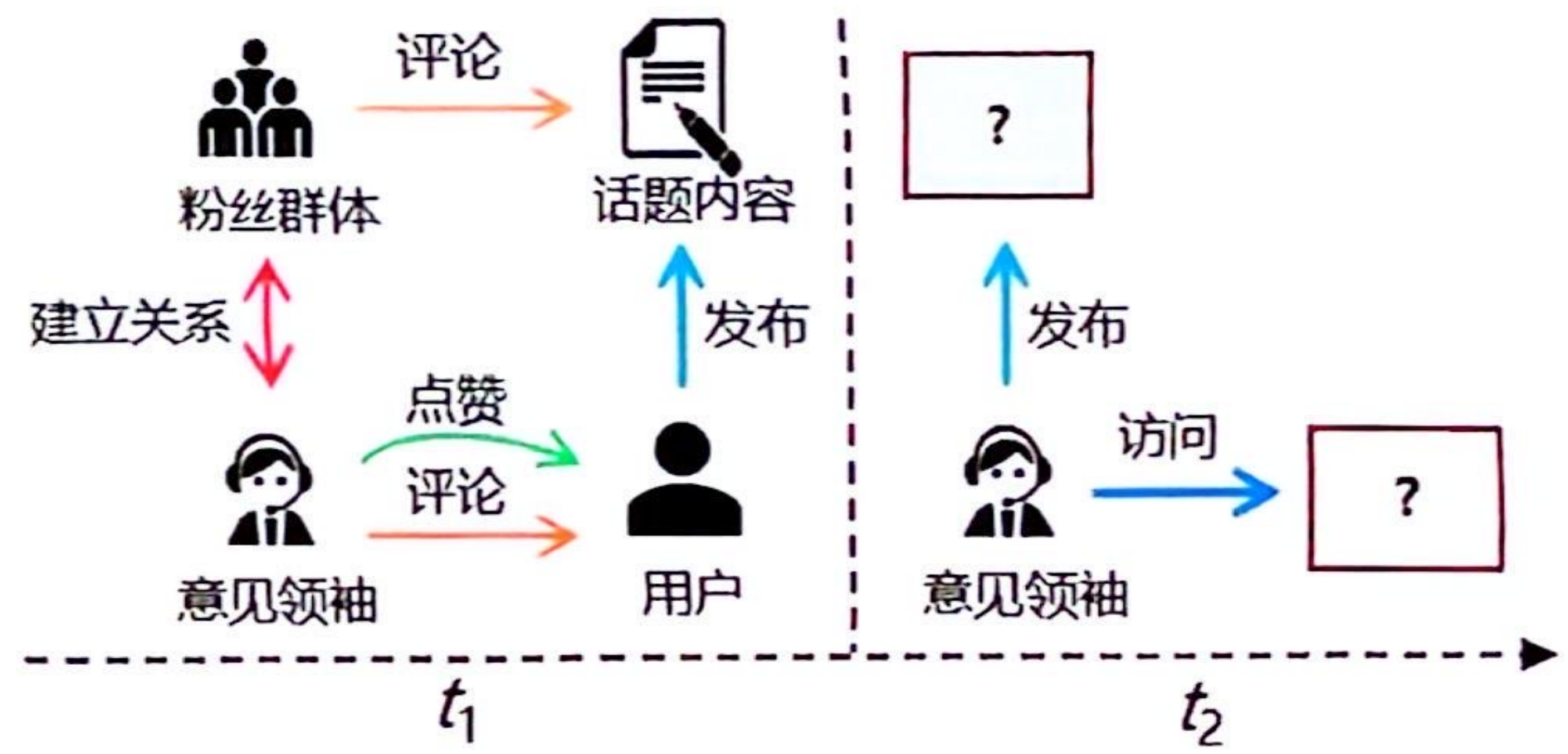
线索发现与预警



类案推荐

● 时序知识图谱推理

时序知识图谱推理指利用时间信息和图结构，通过深度学习方法预测实体关系的未来变化趋势。**问题:** (意见领袖, 发布, ?, t_2)



知识引导，推理因果，溯源治理

技术挑战

- 司法业务态势动态多变；
- 数据量大、标注数据匮乏；
- 多视角评估时效性要求高。

- 行为发生偶然性强；
- 演化途径和规律未知；
- 多视角信息碎片化。

- 行为主客体关系复杂；
- 知识主体分布不均衡；
- 多通道非共享、非开放。

- 数据多通道多模态；
- 数据存在语义鸿沟；
- 数据和先验知识不完备。

研判事理发展

趋势预测

理解案件因素

因果回溯

分析预测

风险预测

预警办案风险

要素发现

定位办案要素

分布式超量学习、深度卷积神经网络

知识图谱推理、元学习、演化博弈论、图神经网络

协同学习、图卷积神经网络、异质网络挖掘

深度自编码器、异质图神经网络、元学习、对比学习、生成对抗网络

技术应对

司法领域知识增强的法律监督规则挖掘

难点与方法

◆ 异构融合难

- 借助**大模型增强的行业知识快速抽取**，构建语义融合的知识图谱。

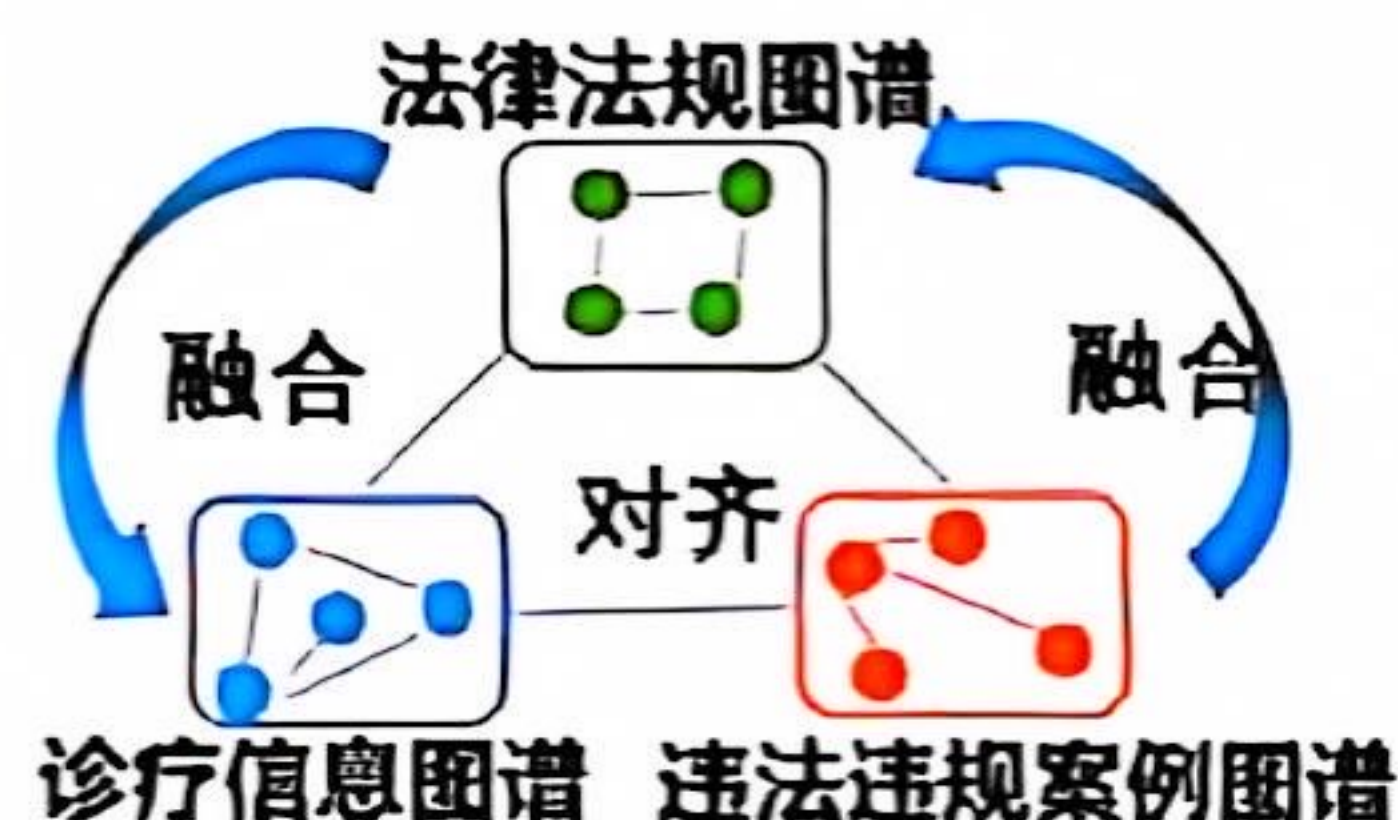
◆ 推理分析难

- 基于图神经网络的**监督要素知识推理**，推理逻辑规则自动发现。

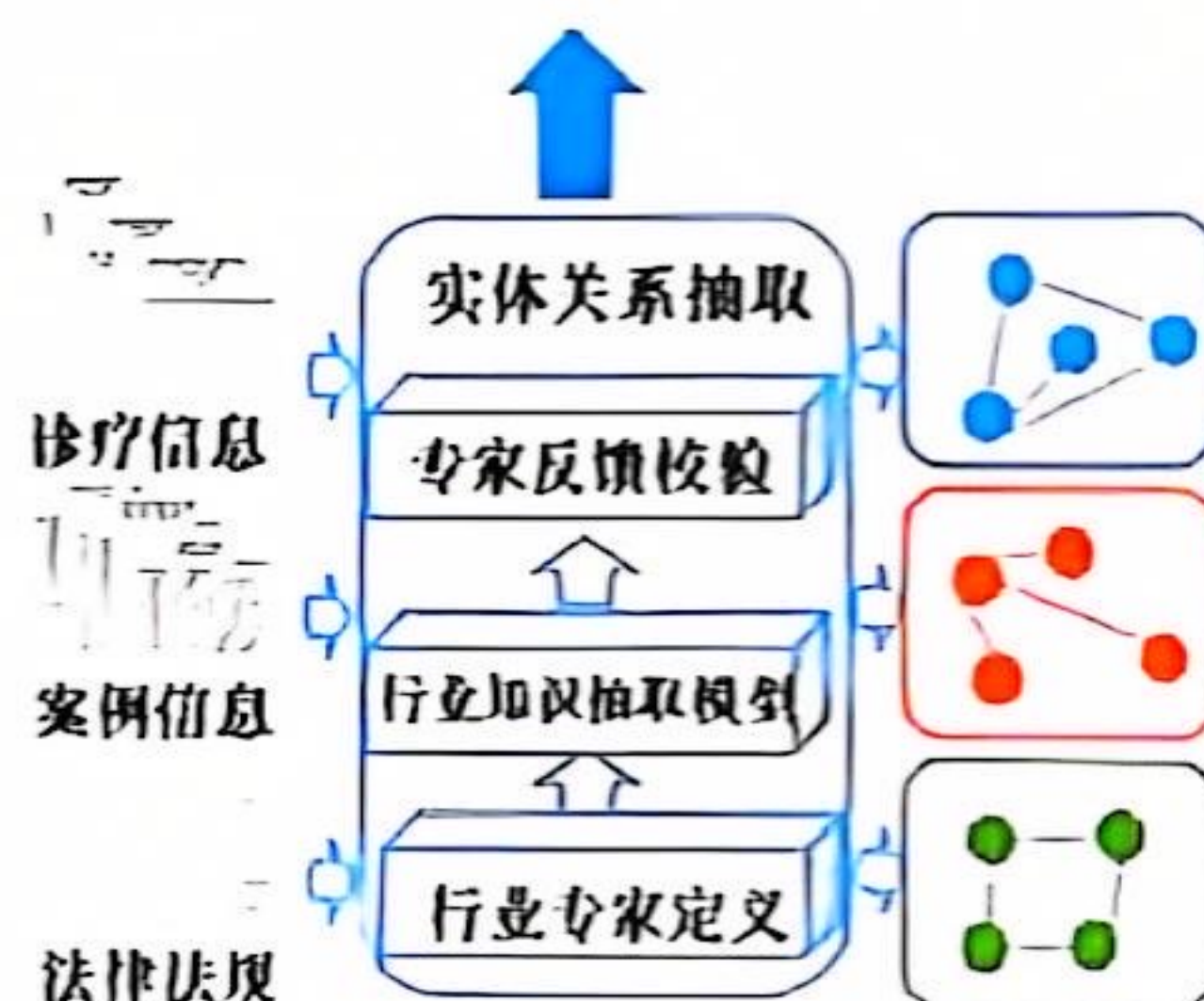
◆ 结果解释难

- 基于**多模态特征的提示工程**，生成场景自适应的可解释法律监督规则。

1. 跨行业知识图谱融合

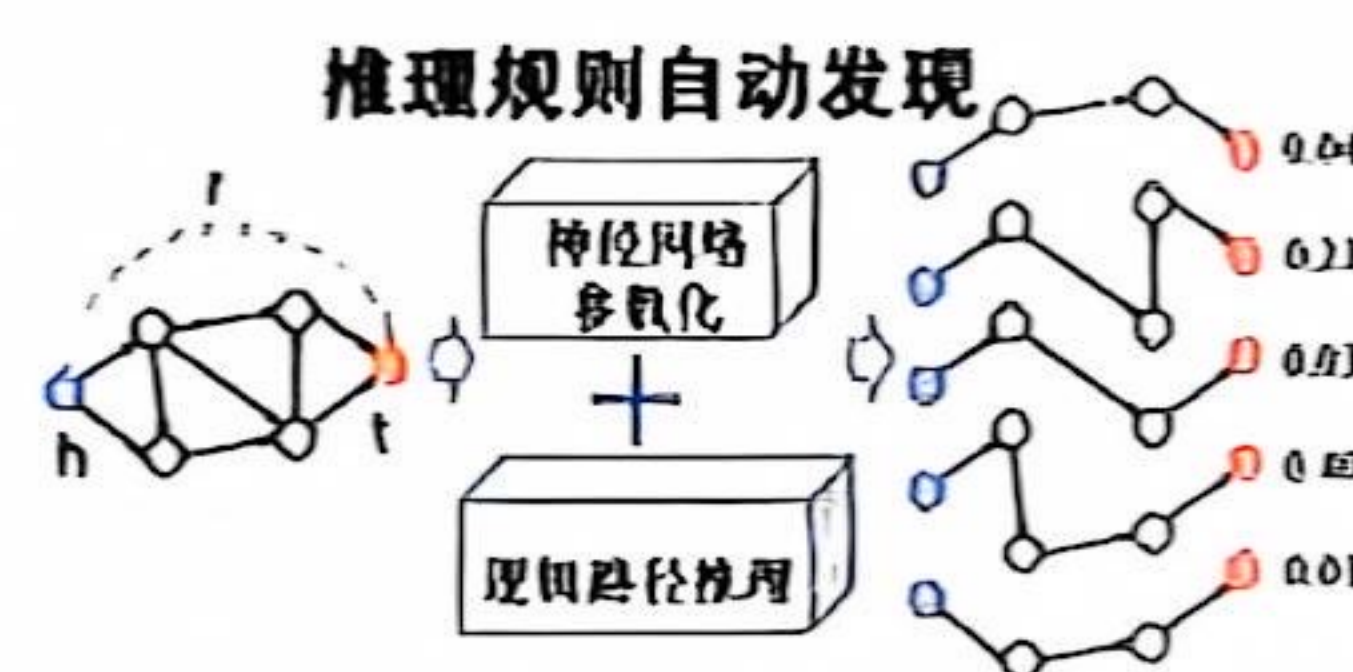


基于深度语义匹配的
跨行业知识融合

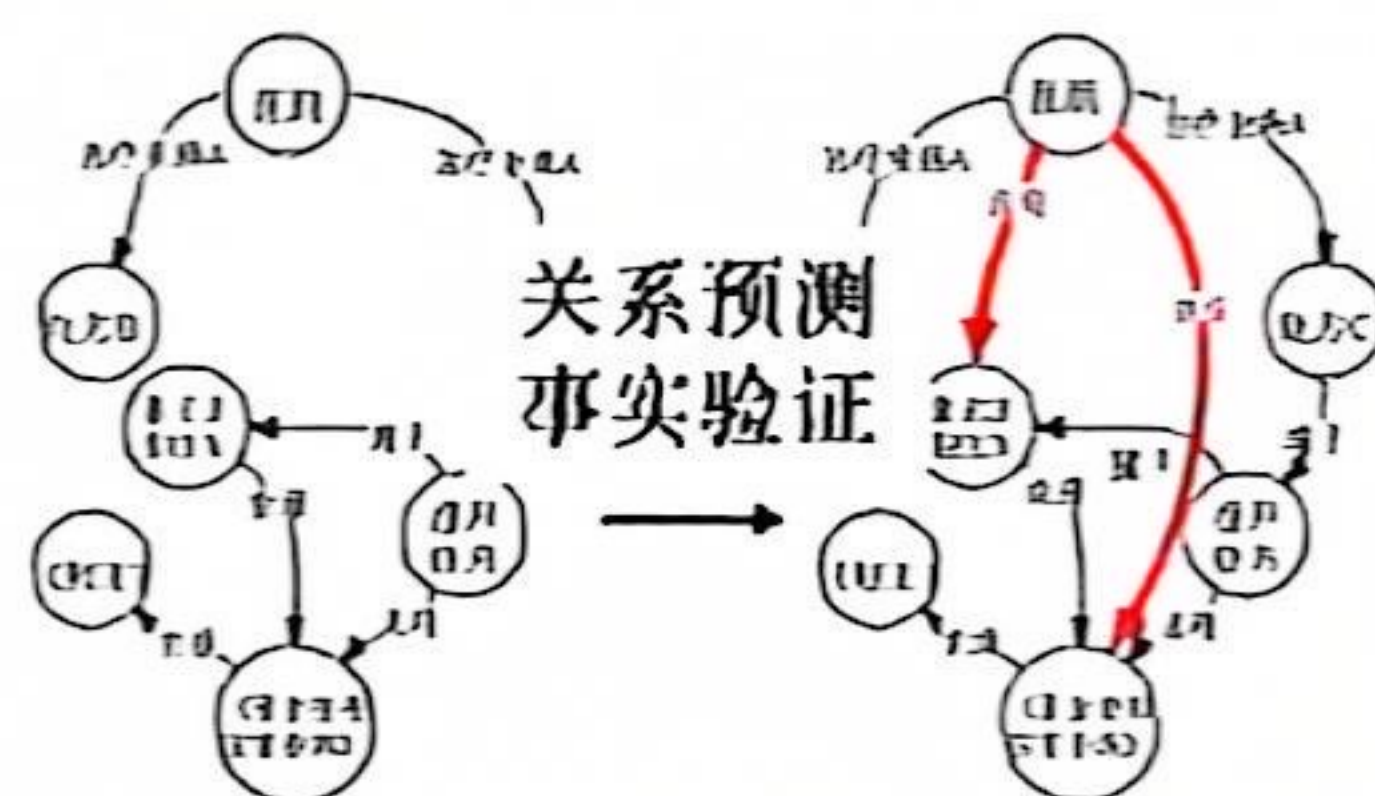


大模型增强的
行业知识快速抽取

2. 监督要素知识推理

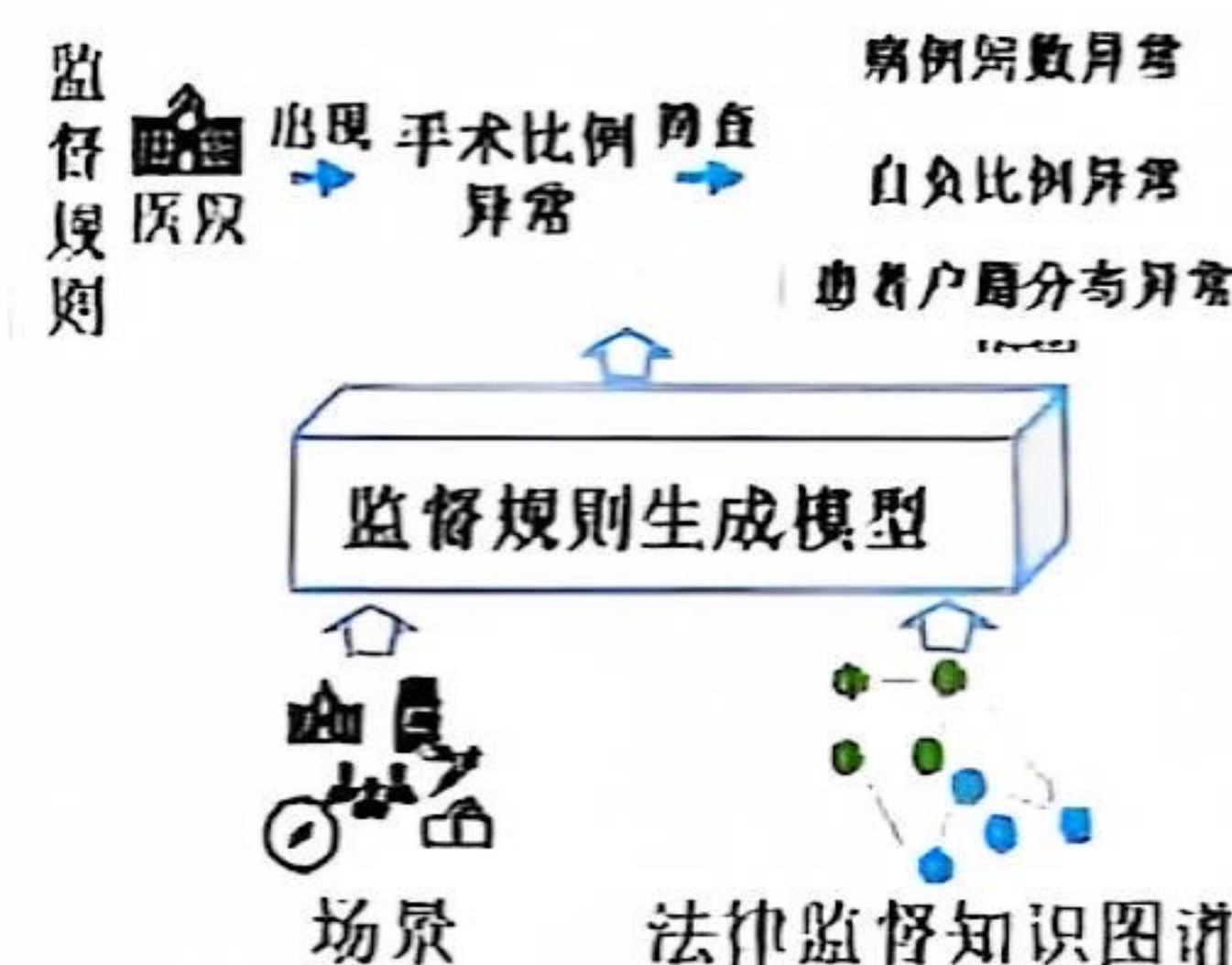


基于深度逻辑归纳的
监督推理规则挖掘

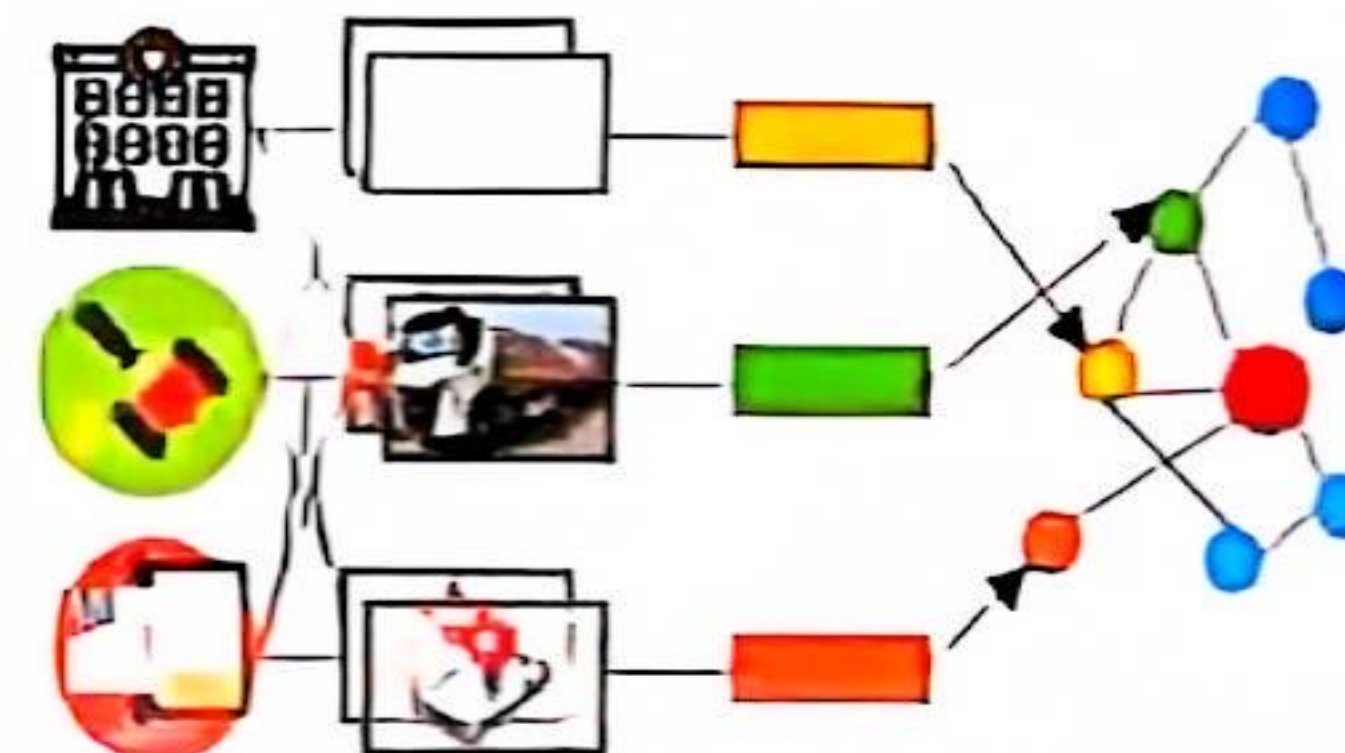


基于图神经网络的
监督图谱知识推理

3. 自适应监督规则生成

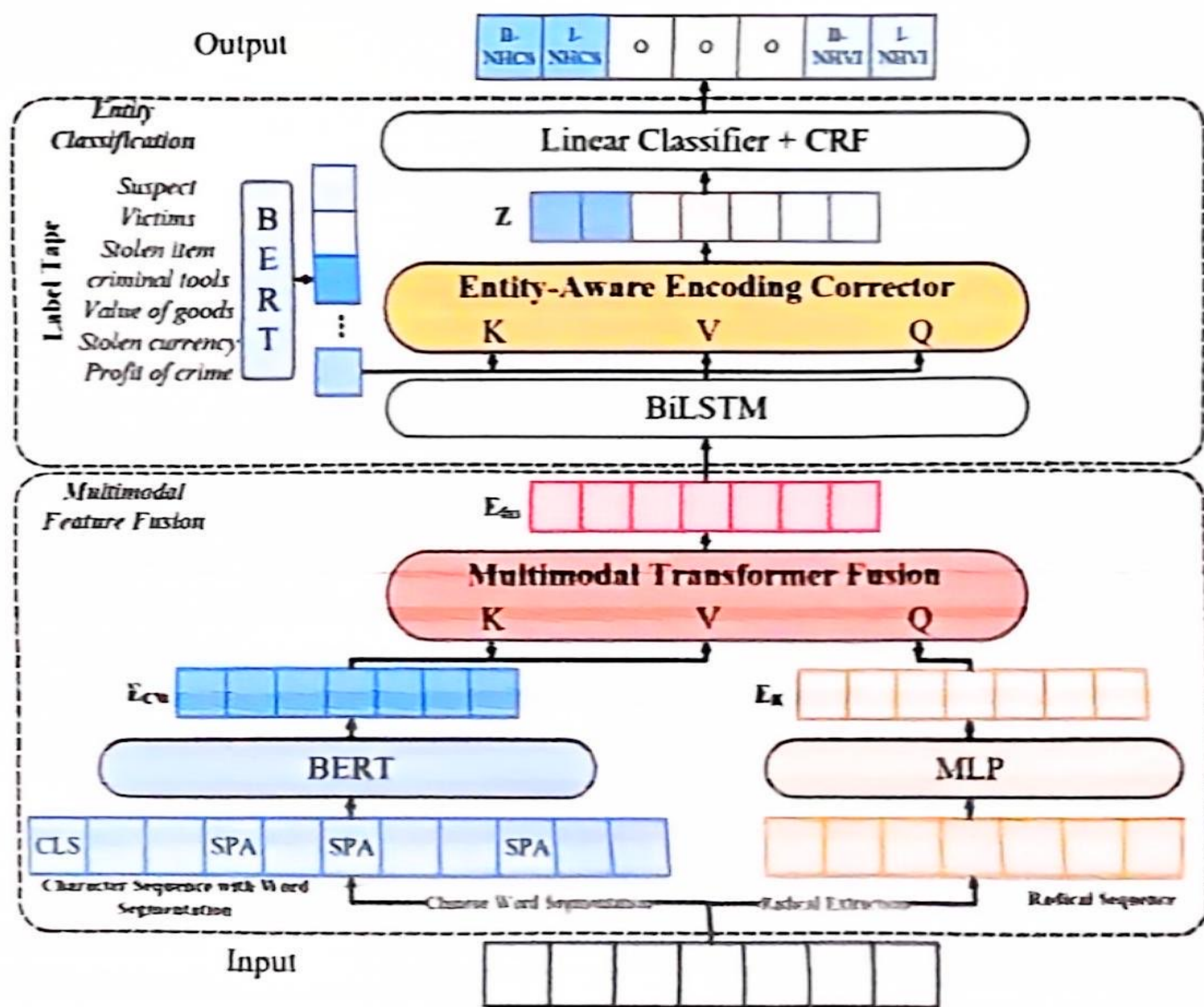


深度预训练增强的
场景自适应监督规则生成



多模态特征的提示工程

基于中文司法领域细粒度命名实体模型



技术方案

- ✓ 多级别特征增强：部首、分词信息结合字嵌入利用注意力机制自适应特征融合，适应司法领域特定词元分布
- ✓ 实体类别细粒度：实体描述引导词元表征贴近相应司法实体

方法	CAIL2021	CAIL2022
RoBERT	88.62	80.45
Chinese-BERT	83.79	78.7
LEBERT	88.47	79.59
W ² NER	89.35	80.12
LegalNER	92.72	85.92



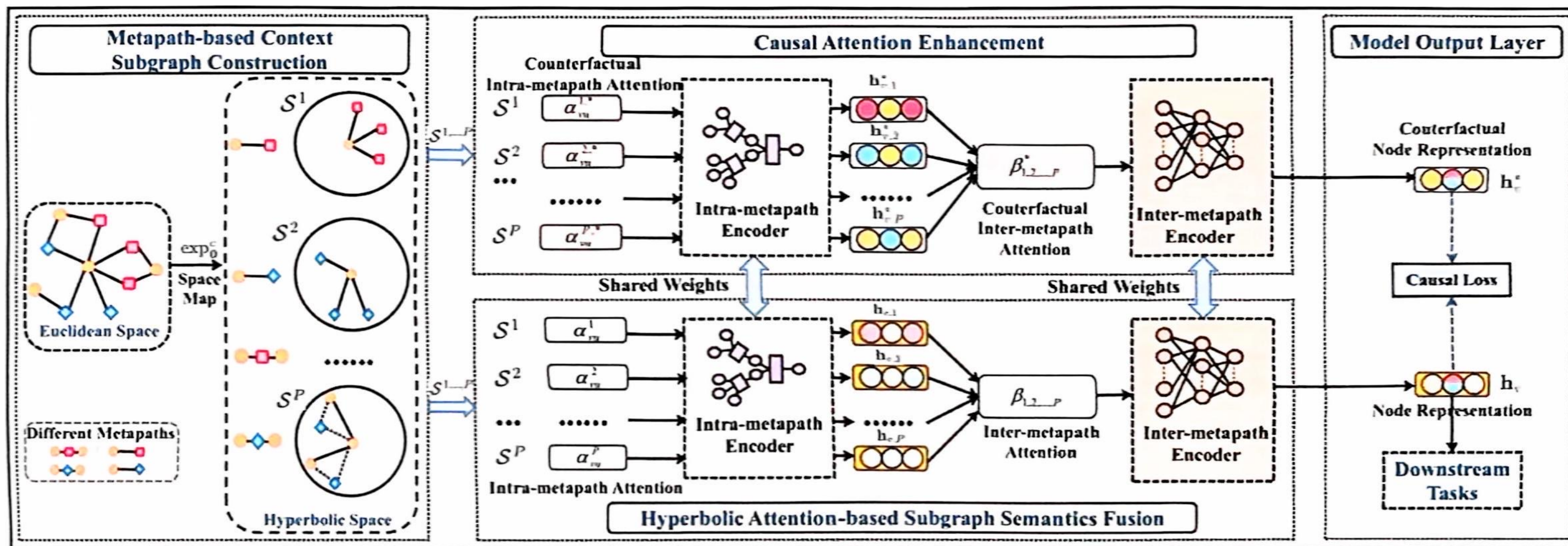
中文司法实体识别模型LegalNER

司法数据集上F1指标

2024IEEE DSC 杰出论文

Li Changqing, Qian Ye, Pan Li, Multimodal Features Enhanced Named Entity Recognition based on Self-Attention Mechanism , 2024 IEEE DSC, Outstanding Paper

因果注意力增强的鲁棒双曲空间异质图表示学习

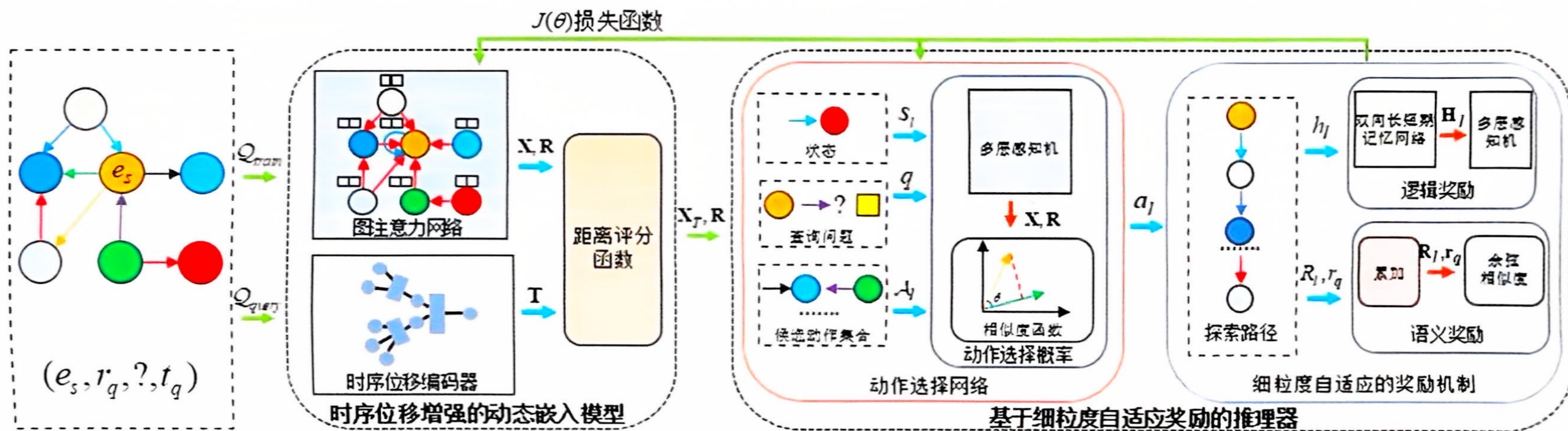


- 针对**语义融合不足**，综合考虑一阶和高阶元路径构建上下文子图，在双曲空间基于注意力进行元路径内和元路径间的信息聚合。
- 针对**虚假语义关联**，通过应用反事实注意力来增强原始的注意力，使注意力学习过程与下游任务对齐，增强模型鲁棒性和泛化性。

Liu Baijia, Zheng Conghui, Pan Li, CATCH: Causal Attention Enhanced Robust Hyperbolic Heterogeneous Graph Representation Learning. *Neural Networks*, 2024

基于细粒度强化学习的可解释时序知识图谱推理

- **研究目标:** 在时序知识图谱上进行可解释的推理。
- **技术方案:** 利用实体之间复杂的**时序依赖关系**来表征实体，从而挖掘事件的时间演变规律；基于实体和关系表征，设计一种**细粒度的自适应奖励机制**，通过深入分析探索路径的逻辑和语义来优化推理器。



医保药品倒卖骗取医保基金类案线索挖掘法律监督模型

■ 技术方案

- ✓ 分析药品**购买频率、种类和数量**，识别异常购药情况
- ✓ 对参保人记录进行**跨医院比对**，推断诊断、处方和药品使用记录一致性

■ 应用成效

- ✓ 2024年“全国检察机关大数据法律监督模型管理平台”完成部署与试用，并基于真实场景数据对模型进行验证。
- ✓ 模型在2023~2024年上海市医保数据上进行验证，涉及医疗机构1,341家，人员8,990人，包括中高风险医保卡5,300张。

