所属项目：面向溯源治理的多元跨数据权属法律监督线索挖掘技术研究
所属专项：社会治理与智慧社会科技支撑（平安中国）
2023 年国家重点研发计划重点专项项目

基于图异常挖掘的欺诈检测在数字检察中的应用
--以医保基金公益诉讼法律监督为例

报告人： 纪文迪
所属课题： 面向行业知识增强的法律监督规则生成
承担单位： 华东政法大学

2024年11月23日

大数据法律监督发展与现状

监督模型

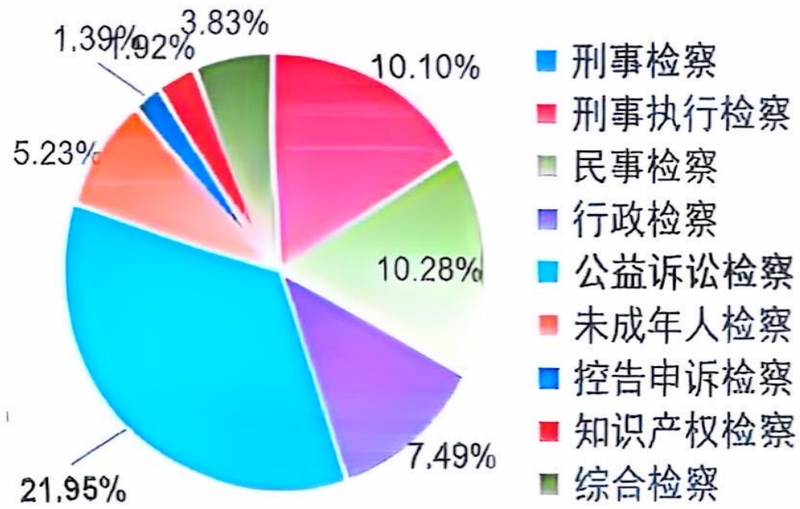
越城区检察院：仅绍兴地区一家医院下属的两个科室，初步查明非法套取医保基金达2000余万元，涉及中老年病患者数千人。

经典案例

2013年1月至2016年8月，刘某甲等三人通过虚假就诊、虚开药方、虚假住院等手段空刷医保卡，非法获取医疗保险金超1亿元。

大数据法律监督在各类检察业务应用情况

(2022 年 6 月)



我国法律监督工作进入数字时代

2022年研发 **大数据监督模型** 800余 个

发现 类案监督线索 20万余条

监督案件 6.1万件

"知识"赋能 智能法律监督

- 2017年6月 最高检印发《检察大数据行动指南》
- 2021年6月 党中央印发《中共中央关于加强新时代检察机关法律监督工作的意见》
- 2021年11月 最高检确定第一批全国检察机关大数据法律监督研发创新基地
- 2022年11月 最高检数字检察办公室成立
- 未来

“类案”引导法律监督模型构建

法律监督场景1：村卫生室医保诈骗

《村卫生室医保诈骗大数据监督模型》--安徽省蒙城县检察院

据统计全国共有村卫生室58.8万个，占医疗机构56.9%，医保诈骗形势严峻，仅2022年全国就追回医保资金188.4亿元。

已在安徽省、广东省投入应用，发现可疑线索五千余条

背景

两起案件：
- 同一个乡镇
- 不同卫生室

是否存在某种关联

监督规则

死亡人员“幽灵报销”：
- 比对医保数据与火化人员数据 → 发现存在已死亡人员仍诊疗报销

小额诊疗处方异常：
- 统计村卫生室小额处方数量 → 个别村卫生室小额处方1000次以上

家庭成员集中就诊：
- 比对医保数据与户籍信息 → 家庭成员在同时就诊、长期连续就诊

医保报销总额极大：
- 统计村卫生室医保报销总额 → 个别卫生室明显高于全县99.6%的卫生室

异常库存虚增：
- 对比药品报销和采购数据→存在报销金额大于采购入库金额的情况

“类案”引导法律监督模型构建

法律监督场景2：第三方责任医保诈骗

湖北省枣阳市：医保基金先行支付案件法律监督
江苏省连云港市赣榆区：医疗保险基金(外伤类)保护检察监督模型

类案监督场景：交通事故

连云港市赣榆区人民检察院
医疗保险基金（外伤类）保护检察监督模型

数据	报销明细信息、公安机关交通事故案件信息	
监督要素	姓名、身份证号、入院时间、交通事故发生时间	
监督规则	条件1	两组数据身份信息一致
	条件2	医保报销数据中的入院时间和交通事故发生时间在同一个时间段
	推理规则	如果同时满足条件1条件2，则存在被他人撞伤走医保报销

比对就医时间与交通事故时间

类案监督场景：人身损害赔偿纠纷

浙江省宁波市宁海县人民检察院
民事检察监督‘N+1’系统

数据	医保报销明细信息、案件裁判文书	
监督要素	姓名、身份证号、入院时间、案件发生时间	
监督规则	筛选	设置‘医疗费’‘伤残等级’等关键词，筛选出人身损害赔偿纠纷案件裁判文书
	条件1	两组数据身份信息一致
	条件2	医保报销数据中的入院时间和纠纷发生时间在同一个时间段
	推理规则	如果同时满足条件1条件2，则存在人身损害赔偿纠纷走医保报销

比对就医时间与纠纷发生时间

医保基金法律监督面临的挑战

有组织的团伙欺诈日益猖獗，这种欺诈行为包含多名患者和医疗服务提供者的联合欺骗活动，给现有的检测方法带来了巨大挑战。因此，团伙医保欺诈检测旨在通过构建并分析医疗保险索赔关系图，检测个体和团伙级别的欺诈索赔，支持医保基金的法律监督。

- 典型案例：2020年年初，陈某甲、陈某乙通过指使参保人至多家医院超量开药、超范围用药，倒卖医保药品，骗取医保基金，涉案金额百余万元。（人民法院案例库）

陈某甲、陈某乙等诈骗、掩饰、隐瞒犯罪所得案

——[illegible]

关键词

[illegible]

基本案情

2020年年初，[illegible]

(2023) 沪0116刑初80号

以非法占有为目的，指使、授意上游骗保人员多开药品，骗取医保基金百余万元

主要挑战

- ✓ 多元化的欺诈骗保手段
- ✓ 现有监管模式的局限性：
  - ✓ 依赖于随机抽查和人工审计
  - ✓ 依靠群众举报和案件线索进行排查

大数据与AI赋能的欺诈检测

常见欺诈分类：

重点人员骗保	死亡人员、重点人员
挂床住院	人床分离
分解住院	未按临床出院指标办理出院、入院，规避医保限额
团伙住院	多人同时空多次住院
意外伤害骗保	第三方途径赔付后，重复报销
卡聚集	多卡聚集，刷卡购药套取医保

基本概念：

•保险欺诈是指个体或组织通过虚假或不当的手段从保险系统中获利的行为。

•特别是在医疗保险领域，欺诈行为可能包括虚假索赔、过度医疗或重复提交费用等

任务目标：

•通过分析保险索赔数据中的异常模式，检测和阻止欺诈行为。

•在动态场景中，保险欺诈检测需要适应不断变化的欺诈模式，这一挑战对传统静态检测方法提出了重大考验

保险欺诈检测的方法：

•规则驱动与传统方法：使用统计和机器学习方法分析特定的欺诈模式。

•图神经网络 (GNN)：多关系图建模、预训练图学习。

未来方向：

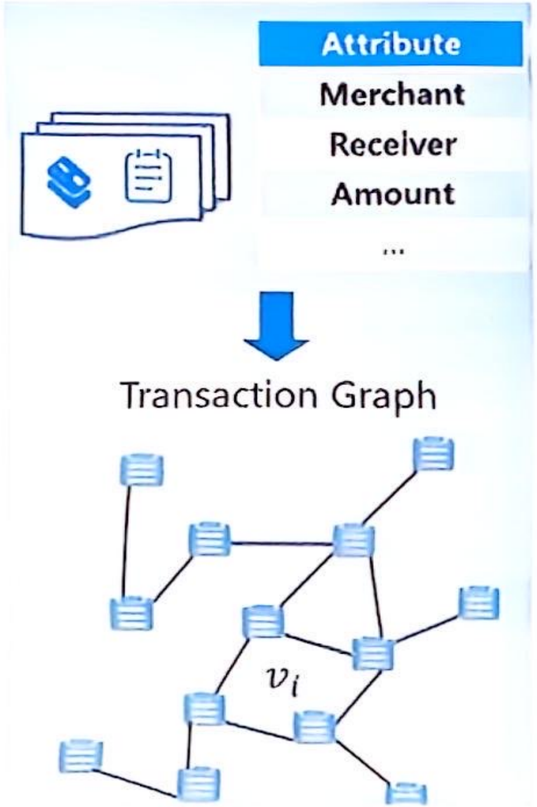
- 深化因果推断，进一步提高模型解释性和可扩展性。
- 发觉有组织的团伙行为。

大数据与AI赋能的欺诈检测

常见欺诈分类：

银行交易欺诈

从交易数据中发现欺诈行为的过程，以减轻因欺诈交易对个人和金融机构造成的损失。



基本概念：

•银行交易欺诈欺诈涉及未经授权的资金使用，目标是识别伪装成合法交易的欺诈行为。

•欺诈者通常伪装成普通用户，通过直接或间接交易规避检测。

任务目标：

•从交易数据中准确检测欺诈行为，同时最小化对合法交易的干扰。

•任务通常建模为一个节点分类问题，其中交易被表示为节点，欺诈行为被标记为异常类别

挑战：

•欺诈行为的隐蔽性：欺诈交易往往隐藏在大量合法交易中，行为模式复杂且多样。

•数据不平衡：欺诈交易占比小，导致模型倾向于预测大多数类别。

•行为变化：用户交易行为随时间变化，称为“概念漂移”

代表性方法：

•基于图的模型：GNN在捕捉复杂交易关系方面表现突出，尤其适用于数据稀疏场景。

•基于时间序列的模型：时间序列模型擅长处理行为变化显著的用户群。

•混合方法：当数据质量差和标签稀缺时，通过集成学习和特征聚合提升检测性能。

大数据与AI赋能的欺诈检测

常见欺诈分类：

重点人员骗保	死亡人员、重点人员
挂床住院	人床分离
分解住院	未按临床出院指标办理出院、入院，规避医保限额
团伙住院	多人同时空多次住院
意外伤害骗保	第三方途径赔付后，重复报销
卡聚集	多卡聚集，刷卡购药套取医保

基本概念：

•保险欺诈是指个体或组织通过虚假或不当的手段从保险系统中获利的行为。

•特别是在医疗保险领域，欺诈行为可能包括虚假索赔、过度医疗或重复提交费用等

任务目标：

•通过分析保险索赔数据中的异常模式，检测和阻止欺诈行为。

•在动态场景中，保险欺诈检测需要适应不断变化的欺诈模式，这一挑战对传统静态检测方法提出了重大考验

保险欺诈检测的方法：

•规则驱动与传统方法：使用统计和机器学习方法分析特定的欺诈模式。

•图神经网络 (GNN)：多关系图建模、预训练图学习。

未来方向：

- 深化因果推断，进一步提高模型解释性和可扩展性。
- 发觉有组织的团伙行为。

数据驱动的异常检测

异常检测算法分类：

- 有监督的异常检查
  - 基于有标签的合法和欺诈数据构建模型
  - 无法识别新的可疑活动
- 无监督的异常检测
  - 无需标注数据
  - 能够识别新的可疑活动
- 半监督的异常检测
  - 用时使用有标注数据和无标注数据
  - 需要少量有标注数据
  - 能够识别新的可疑活动

基于特征的异常检测方法：

- Support Vector Machine (SVM)
- logistic regression (LR)
- k-means clustering
- k-nearest neighbours (kNN)...

基于时序数据的异常检测：

- ARIMA、LSTM、Transformer...

基于图的异常检测：

- 基于结构的图异常检测、基于社区的图异常检测、基于分解的图异常检测、基于概率的图异常检测...

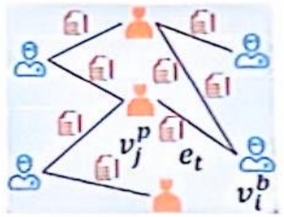
基于图异常挖掘的医保欺诈检测

图数据可以捕捉到实体之间的复杂关系。

图异常挖掘：从复杂的关系网络中识别异常模式的机器学习算法。

图构建

节点：参保人
边：关联行为

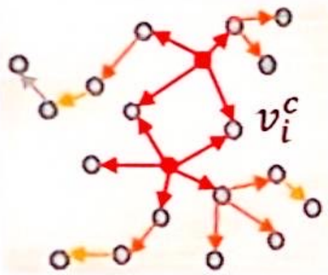


图转化

参保人的就医记录、购药行为、住院记录等数据

欺诈检测

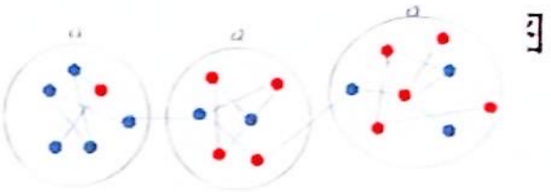
无监督离群节点检测（UNOD）



结构性异常节点：与不同社区有异常链接的节点 → 患者在不同医院频繁住院或开药

情境性异常节点：结构上与周围节点一致，但属性上存在显著差异的节点 → 医生在短时间内开出大量处方

团伙画像



团伙群体属性画像、行为画像

团伙个体属性画像、行为画像

风险与未来挑战

图异常挖掘算法的技术局限性：

- 图异常挖掘技术依赖于高质量的数据。
- 技术难点在于取证的困难。

隐私与数据安全： 图异常挖掘技术依赖于大量的个人医疗数据和行为数据，这些数据的采集、存储和分析过程都可能涉及到个人隐私的泄露风险。

- 数据管理和访问控制机制
- 采用数据脱敏、匿名化等隐私保护技术
- 数据共享平台构建

法律与伦理问题

- 数据滥用的风险
- 图异常挖掘技术的应用可能对商业秘密产生影响
- 偏见和不公正问题