

探索与实践

基于知识图谱与加权贝叶斯机制的 刑侦推断模型

李 波¹, 章 勇^{1,2}, 胡誉骞¹

(1. 华中师范大学 数学与统计学学院, 湖北 武汉 430079; 2. 合肥第六中学, 安徽 合肥 230061)

摘 要: 随着信息化和数字化的高速发展, 各类人工智能方法在现代公共安全领域得到充分应用, 其中知识图谱以其强大的知识表征能力和数学建模能力, 大大提高了智能决策的可解释性, 成为领域内研究和应用的热点. 一直以来, 基于贝叶斯机制的推断模型是公安情报与刑侦分析的重要手段. 本文将公安知识图谱引入传统贝叶斯模型, 在充分考虑参数耦合性的基础上, 对后验概率进行加权更新, 有效提高了推断的准确性. 本文回避敏感的真实数据, 以《红楼梦》中“绣春囊”一案为例完成实证分析, 科学地揭秘了经典悬案的真相. 本文提出的基于知识图谱的自适应加权方案有效提高了贝叶斯模型的效率, 是人工智能助力智慧公安的一种有益尝试.

关键词: 知识图谱; 贝叶斯模型; 关系网络权重; 自适应加权

中图分类号: O29

文献标志码: A

文章编号: 2095-3070(2024)03-0015-11

DOI: 10.19943/j.2095-3070.jmmia.2024.03.02

0 引言

国家高度重视公安科技创新, 强调“坚持政治治警、改革强警、科技兴警、从严治警, 履行好党和人民赋予的新时代职责使命”; “要把大数据作为推动公安工作创新发展的大引擎、培育战斗力生成新的增长点, 全面助推公安工作质量变革、效率变革、动力变革”^[1]. 近年来, 我国公安工作的科技支撑得到飞速提升, 各级公安机关充分利用现代信息技术和资源, 极大地推动了公安工作的数字化和智慧化, 特别地, 数学模型成为此进程中的重要基石.

贝叶斯理论最常见的形式是贝叶斯定理^[2], 它的本质思想是根据先验信息提高计算后验概率的准确性. 贝叶斯理论遵循人类思维和科学分析的逻辑, 具有广泛的应用场景, 在公共安全领域也有丰富的理论和实践案例. 如定量犯罪学^[3]、证据与嫌疑的关联估计^[4]、失事飞机发现^[5-6]、犯罪地点分布预测^[7]、连环犯罪行为关联^[8]、犯罪在场概率估计^[9]、青少年犯罪测试^[10]和嫌疑人摸排^[11]等. 在情报分析中, 贝叶斯定理结合掌握的片段证据信息, 确定结果的可能性, 通过概率估算将证据信息转为量化、高质量的决策依据, 提高警务效率与准确性^[4]. 陈雨楠等^[12]基于贝叶斯理论对公安情报问题进行预测, 陈亮等^[13]总结了公安情报中的贝叶斯推理机制, 并进一步挖掘线索进而解决公共安全问题的.

传统的贝叶斯推断对参数进行估计时, 要求参数具有互斥性, 忽略了参数之间的相关关系, 在一

收稿日期: 2024-04-16

基金项目: 国家自然科学基金面上项目(62377019); 湖北省高校省级教学研究项目(2022083); 武汉市科技计划知识创新专项项目(2022010801010273)

通讯作者: 胡誉骞, E-mail: wujyuhin@163.com

引用格式: 李波, 章勇, 胡誉骞. 基于知识图谱与加权贝叶斯机制的刑侦推断模型[J]. 数学建模及其应用, 2024, 13(3): 15-25.

LI B, ZHANG Y, HU Y Q. Criminal investigation inference model based on knowledge graph and weighted Bayesian mechanism(in Chinese)[J]. Mathematical Modeling and Its Applications, 2024, 13(3): 15-25.

定程度上存在不合理性,例如在犯罪团伙拐卖妇女的案件中,团伙的亲属关系对共同犯罪的影响呈显著正相关关系^[14]。此外,在简单离散情况下,后验概率可能具有很大的同质性,导致贝叶斯模型难以给出合理、精准的决策^[15]。

知识图谱作为知识结构化的表现形式,具有强大的知识建模和表征能力^[16],在各种商业和科学领域有多种应用,如推荐、检索和推理等^[17],日益成为人工智能(AI)领域备受关注的重要一环。随着知识图谱在公共安全领域的推广,公安情报知识图谱概念逐渐衍生,如周鹏等^[18]构建家族族谱网络,为家族团伙类型违法犯罪研究提供辅助;樊舒^[19]首次总结了公安情报知识的数据规则与图谱框架。本文利用公安知识图谱来解析复杂系统背后关系,并在贝叶斯公式的理论基础上,通过加权调整后验概率的计算,将各种相关性嵌入其中,以进一步优化后验概率的估计。本文提出了一种基于知识图谱的自适应加权方案,该方案有效提升了贝叶斯模型的应用效率,从而为刑事侦查工作提供了有力的支持。

1 基于知识图谱的关系加权贝叶斯模型构建

本节首先介绍公安情报知识图谱,根据图谱建立关系加权贝叶斯模型,并给出保证算法合理性的相关理论性质。

1.1 公安情报知识图谱构建

本文采用樊舒提出的公安情报知识图谱构建方法。根据数据类型明确的原则,本体一般包括事件、行为、时间、地点、人物、组织机构和物品7类,如图1^[20]所示。而本体具有相应属性,例如,人物本体具有年龄、性别、住址和家庭关系等个人信息。实际工作中,根据不同应用场景需求,知识图谱的搭建会有相应不同的、更为专业细致的数学建模。

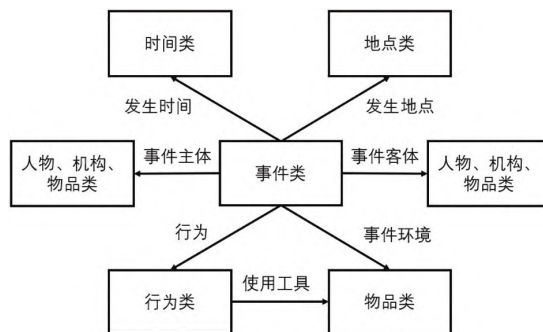


图1 公安情报知识图谱构建原则

1.2 贝叶斯模型机制推演

结合已有的贝叶斯模型研究与公安情报知识图谱,本文建立基于知识图谱的关系加权贝叶斯模型,其计算流程如图2所示。

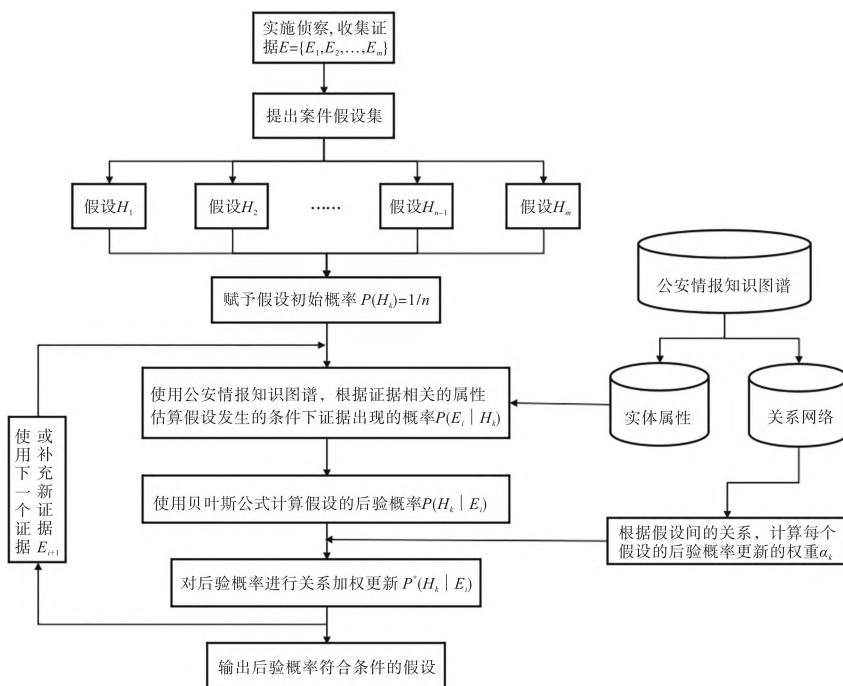


图2 基于知识图谱的关系加权贝叶斯模型计算流程图

首先, 根据事件假设的先验概率与条件概率, 计算在发现证据的条件下各假设发生的后验概率; 然后, 根据关系网络计算每个假设的权重, 对后验概率加权更新; 最后, 将更新的后验概率作为先验概率重新代入模型, 求下一个证据的后验概率; 以此类推, 考虑所有证据后, 得到最终的贝叶斯概率分布, 并结合相应规则给出推断结果.

步骤 1 建立假设集与证据链

根据案件结果, 建立案件的假设事件集 $H = \{H_1, H_2, \dots, H_n\}$, H_k 在公安领域中一般记为第 k 位人物涉嫌违法犯罪. 根据案件分析建立独立证据链 $E = \{E_1, E_2, \dots, E_m\}$, E_i 记为证据链的第 i 个证据, 且证据链可随时间增长.

步骤 2 给定初始概率

在初始无证据的情况下, 假设事件集中每一个假设的初始概率相等, 即 $P(H_k) = 1/n$.

步骤 3 计算条件概率

条件概率 $P(E_i | H_k)$ 表示“第 k 个假设成立下, 第 i 个证据发生的概率”, 因此条件概率与假设事件 H_k 有关. 本文提出一种计算的思路: 如果证据 E_i 相关的属性与案件呈正相关关系, 对应属性数值为 a_k , 则条件概率的计算使用公式(1); 如果证据 E_i 相关的属性与案件呈负相关关系, 对应属性数值为 b_k , 则条件概率的计算使用公式(2).

$$P(E_i | H_k) = a_k / \sum_{j=1}^n a_j, \quad k = 1, 2, \dots, n; \quad (1)$$

$$P(E_i | H_k) = \frac{1}{b_k} / \sum_{j=1}^n \frac{1}{b_j}, \quad k = 1, 2, \dots, n. \quad (2)$$

步骤 4 计算后验概率

通过贝叶斯公式, 计算给定证据 E_i 下假设 H_k 成立的后验概率:

$$P(H_k | E_i) = \frac{P(H_k, E_i)}{P(E_i)} = \frac{P(H_k | E_{i-1}) P(E_i | H_k)}{\sum_{l=1}^n P(H_l | E_{i-1}) P(E_i | H_l)}, \quad (3)$$

当 $i=1$ 时, $P(H_k | E_{i-1}) = P(H_k)$.

步骤 5 计算后验概率更新权重

不同实体之间关系性质的不同, 其内在的合作和竞争类型与程度也有所差异. 一般而言, 两个个体如果关系亲密、接触频繁、年龄接近等, 则他们一致性程度较高. 因此权重更新重点考虑两个方面: 一是个体的概率分布; 二是个体之间的关系密切程度.

记顶点集 $V(G) = \{v_1, v_2, \dots, v_n\}$ 为节点集合, 边集 $E(G) = \{e_1, e_2, \dots, e_h\}$ 为人与人之间的关系集, n 为人数, h 为图谱的关系数量. 指定第 k 人, 与他有关的人物集合记为 $V_k = \{v_1^{(k)}, v_2^{(k)}, \dots, v_{n_k}^{(k)}\}$, 相关的边集合为 $E_k = \{e_1^{(k)}, e_2^{(k)}, \dots, e_{n_k}^{(k)}\}$, 设集合 V_k 和 E_k 对应的索引指标集合为 I_k . 根据关系的亲密程度、接触频率、重要性等赋予关系权重 q , 则后验概率 $P(H_k | E_i)$ 的概率更新权重为:

$$\alpha_k = \frac{\sum_{i \in I_k} q_i^{(k)} p_i^{(k)} + q^{(k)} p^{(k)}}{\sum_{i \in I_k} q_i^{(k)} + q^{(k)}}, \quad (4)$$

其中: $q_i^{(k)}$ 为第 k 个人与第 i 个人的关系权重; $p_i^{(k)}$ 为与第 k 个人有关系的第 i 个人的后验概率; $p^{(k)}$ 为第 k 个人的后验概率; $q^{(k)}$ 为第 k 个人与自己的关系权重.

步骤 6 关系加权更新后验概率

根据权重 α_k 对贝叶斯后验概率进行加权更新. 第 k 个人的基于关系的贝叶斯概率为:

$$p_{ki}^* = P^*(H_k | E_i) = \frac{\alpha_k P(H_k | E_i)}{\sum_{m=1}^n \alpha_m P(H_m | E_i)} = \beta_k P(H_k | E_i). \quad (5)$$

步骤 7 根据证据更新概率

将下一个证据或者新获取的证据纳入, 执行步骤 3、4、5、6, 对每个假设发生的概率进行更新.

1.3 模型的推论

推导两个推论用以保证本文算法的正确性. 推论 1 保证独立证据的选取次序不影响后验概率推断; 推论 2 保证引入关系权重后独立证据的选取次序不影响后验概率推断.

推论 1 贝叶斯公式后验概率与独立证据链 $\{E_1, E_2, \dots, E_m\}$ 的发生顺序无关.

证明: 对任意给定的独立证据 E_i 和 E_j , 第 k 个假设推理 H_k 的概率为:

$$\begin{aligned} P(H_k | E_i, E_j) &= \frac{P(H_k, E_i, E_j)}{P(E_i, E_j)} = \frac{P(H_k, E_i)P(E_j | H_k, E_i)}{\sum_{l=1}^n P(H_l, E_i)P(E_j | H_l, E_i)} \\ &= \frac{P(E_i)P(H_k | E_i)P(E_j | H_k, E_i)}{\sum_{l=1}^n P(E_i)P(H_l | E_i)P(E_j | H_l, E_i)} \\ &= \frac{P(H_k | E_i)P(E_j | H_k)}{\sum_{l=1}^n P(H_l | E_i)P(E_j | H_l)}. \end{aligned}$$

因此, 在独立证据链 $E = \{E_1, E_2, \dots, E_m\}$ 下, 任意证据 E_i 计算的后验概率能作为另一个证据 E_j 的先验概率, 因此贝叶斯公式的后验概率与独立证据链 E 的发生顺序无关.

推论 2 基于关系的贝叶斯后验概率与独立证据链 $\{E_1, E_2, \dots, E_m\}$ 的发生顺序无关.

证明: 任意给定的 E_i 与 E_j , 第 k 个假设的先验概率 $P(H_k) = p_k = 1/n$, $k = 1, 2, \dots, n$, 条件概率 $P(E_i | H_k) = m_{ik}$, $P(E_j | H_k) = m_{jk}$, 记后验概率 $P(H_k | E_i) = p_{ki}$. 在证据链顺序为 E_i, E_j 时, 假设 H_k 在 E_i 情况下的后验概率与基于关系的贝叶斯概率分别为

$$\begin{aligned} p_{ki} &= \frac{P(H_k)P(E_i | H_k)}{\sum_{f=1}^n P(H_f)P(E_i | H_f)} = \frac{p_k m_{ik}}{\sum_{f=1}^n p_f m_{if}}, \\ p_{ki}^* &= \frac{\alpha_k p_{ki}}{\sum_{m=1}^n \alpha_m p_{mi}} = \frac{\alpha_k p_k m_{ik}}{\sum_{m=1}^n \alpha_m p_m m_{im}}. \end{aligned}$$

在 E_j 情况下, 将 p_{ki}^* 作为先验概率, 则贝叶斯概率与基于关系的贝叶斯概率分别为

$$\begin{aligned} p_{kj} &= \frac{p_{ki}^* m_{jk}}{\sum_{f=1}^n p_{fi}^* m_{jf}} = \frac{(\alpha_k p_k m_{ik}) m_{jk}}{\sum_{f=1}^n (\alpha_f p_f m_{if}) m_{jf}}, \\ p_{kj}^* &= \frac{\alpha_k p_{kj}}{\sum_{m=1}^n \alpha_m p_{mj}} = \frac{\alpha_k^2 p_k m_{ik} m_{jk}}{\sum_{m=1}^n \alpha_m^2 p_m m_{im} m_{jm}}. \end{aligned}$$

在证据链顺序为 E_j, E_i 的情况时, 同理推导得到最后基于关系的贝叶斯概率 $p_{ki}^* = p_{kj}^*$. 在两种顺序下的基于关系的贝叶斯概率相等, 使用数学归纳法推广可得到结论: 最终后验概率与独立证据链 $E = \{E_1, E_2, \dots, E_m\}$ 的发生顺序无关.

2 案例应用与分析

本节首先以《红楼梦》为背景构建知识图谱, 并结合图谱与加权贝叶斯模型对“绣春囊”案件建模, 最后进行实例分析, 科学地揭秘该经典悬案的真相. 其中关系图谱中的概率可视化工作使用 neo4j 图数据库完成.

2.1 问题提出

由于真实公安数据涉及数据隐私和安全问题, 本文以《红楼梦》著名悬案“绣春囊”案为例完成实

证分析。“绣春囊”疑案出自书中第七十三回“痴丫头误拾绣春囊，懦小姐不问累金凤”。故事如下。贾母的丫鬟捡到一个绣春囊，而绣春囊在大观园中是不雅之物，有损贾家声誉，因此导致贾府内部查抄大观园。大丫鬟司棋、入画、晴雯被众人搜出男子藏物，最终晴雯、司棋、芳官、入画含冤被逐^[20-21]。绣春囊究竟“归罪”于谁是多年来红学家争论的问题，结合文中暗示与信物推断，绣春囊是潘又安送给秦司棋的，因此该案件最大嫌疑人正是秦司棋^[22-23]。

2.2 《红楼梦》社会关系知识图谱框架构建

根据公安情报知识图谱构建原则，本文对《红楼梦》中的人物及其社会联系构建如图 3 的知识图谱数据建模范式，结合案发时收集的证据链条进行贝叶斯推理。

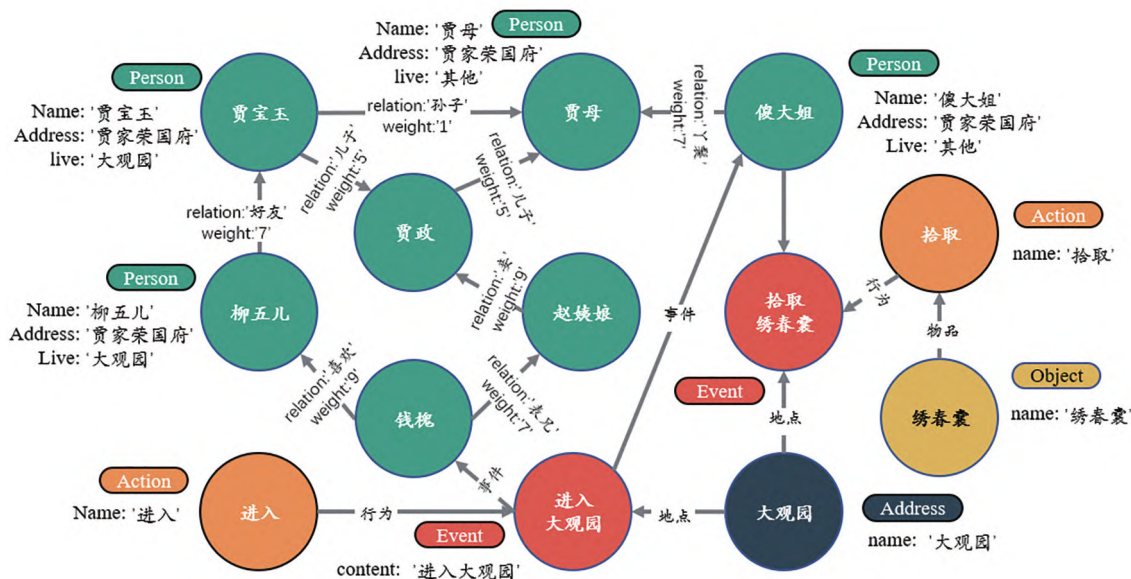


图 3 《红楼梦》知识图谱数据建模范式

2.3 基于红楼梦人物信息的关系加权贝叶斯模型

结合《红楼梦》知识图谱，应用已建立的关系加权贝叶斯模型，给出以下步骤。

步骤 1 建立假设集与证据链

给定假设事件集 $H = \{H_1, H_2, \dots, H_n\}$, H_j 表示绣春囊属于第 j 个人。给定目前搜集到的证据链 $E = \{E_1, E_2, E_3\}$, E_1 表示“失窃物是绣春囊”， E_2 表示“绣春囊丢失地点为大观园后山”， E_3 表示“绣春囊价值低”。

步骤 2 给定初始概率

为每个假设赋予初始概率 $P(H_j) = 1/n$ 。

步骤 3 计算条件概率

证据 E_1 为“失窃物是绣春囊”，此物与男女关系密切，则与该证据相关的信息是实体是否具有亲密关系，此类关系包含丈夫、妻子、二房和陪房丫环等。设 a_j 表示第 j 个人是否具有亲密关系，则其表达式为：

$$a_j = \begin{cases} r_1, & \text{有亲密关系,} \\ r_2, & \text{其他,} \end{cases} \quad (6)$$

其中, $r_1, r_2 \in [0, 1]$, $r_2 < r_1$, $j = 1, 2, \dots, n$ 。

a_j 不取示性函数是为保证不具有亲密关系的人仍具有较低可能性拥有绣春囊。 a_j 与 E_1 呈正相关，则条件概率为：

$$P(E_1 | H_j) = a_j / \sum_{k=1}^n a_k, \quad j = 1, 2, \dots, n. \quad (7)$$

证据 E_2 为“绣春囊丢失地点为大观园后山”，与该证据相关的信息是实体是否住在大观园。设 b_j 为第 j 个人是否去过大观园，其表达式为

$$b_j = \begin{cases} r_1, & \text{去过大观园,} \\ r_2, & \text{其他,} \end{cases} \quad (8)$$

其中, $r_1, r_2 \in [0, 1]$, $r_2 < r_1$, $j = 1, 2, \dots, n$. 而 b_j 与 E_2 呈正相关, 则条件概率为

$$P(E_2 | H_j) = b_j / \sum_{k=1}^n b_k, j = 1, 2, \dots, n. \quad (9)$$

步骤 4 利用公式(3)计算后验概率

步骤 5 计算后验概率更新权重

案例丢失物品是“见不得人”的绣春囊, 因此考虑情人间无意丢失的可能性. 根据关系亲密程度设计权重, 设第 k 个人的关系指标集中的第 i 种关系的权重为 $q_i^{(k)}$, 权重见表 1. 最终根据公式(4)计算后验概率更新权重.

表 1 不同关系的亲密程度权重

关系	权重	关系	权重	关系	权重
自己	10	姐妹	7	儿媳	3
女友	9	养父	5	侄女	3
男友	9	弟媳	5	姨妈	3
妃	9	父亲	5	婆婆	3
续弦	9	乳母	5	女婿	3
妾	9	儿子	5	姑母	3
二房	9	女儿	5	舅舅	3
丈夫	9	奴仆	5	邻居	1
暧昧	9	养子	5	施主	1
相好	9	姐夫	5	奶奶	1
喜欢	9	干娘	5	玄孙	1
妻子	9	干儿子	5	老亲	1
丫环	7	被扶养	5	伙计	1
朋友	7	嫂子	5	买办	1
好友	7	母亲	5	重孙媳妇	1
陪房丫环	7	乳父	5	小厮	1
兄弟	7	门生	3	祖父	1
弟弟	7	岳母	3	外孙女	1
主人	7	岳父	3	爷爷	1
表哥	7	侄子	3	孙子	1
妹妹	7	老师	3	外祖母	1
表兄	7	徒弟	3	臣	1
雇主	7	伯父	3	孙女	1

注: 与自己的权重是最大值, 男女关系权重赋值为 9, 同辈、丫环此类关系赋值为 7, 直系长辈、直系晚辈此类关系赋值为 5, 旁系长辈、旁系晚辈赋值为 3, 辈分高于或低于两层、非亲戚关系等关系赋值为 1.

步骤 6 利用公式(5)计算关系加权后的后验概率

2.4 贝叶斯模型与基于关系的贝叶斯模型结果对比

在传统贝叶斯模型中, 由于条件概率是简单离散分布, 因此经过贝叶斯公式对两个证据更新两次后, 有 8 个个体同时达到最大概率, 其实质是个体同时满足所有证据; 达到第二大概率的个体有 127

个实体,其实质是个体满足其中一个证据;达到第三大概率的个体有 269 个,其实质是个体没有满足任意一个证据,因此分为 3 个排名区间。

在基于图谱的关系加权贝叶斯模型中,由于关系加权贝叶斯模型是结合实体与其社会关系的一致性建模,在贝叶斯公式基础上嵌入了关系网络,因此后验概率具有一定的连续性,且由于关系的双向性,部分相关个体的概率会有耦合提升。经过加权后,原本为前 8 名的概率依然最高,原本为第二大概率的 127 个实体有 92.12% 的个体在第二个排名区间中,原本为第三大概率的 269 个实体中有 96.28% 的个体依然在第三个排名区间中。

图 4 为贝叶斯后验模型结果。大量相同概率的实体同为重点对象,显然优先排查的意义有限,无法确定最可能的嫌疑人。在传统贝叶斯模型中,概率最大且相等的 8 个个体分别为贾迎春、李纨、贾宝玉、薛宝钗、花袭人、柳五儿、香菱和秦司棋。

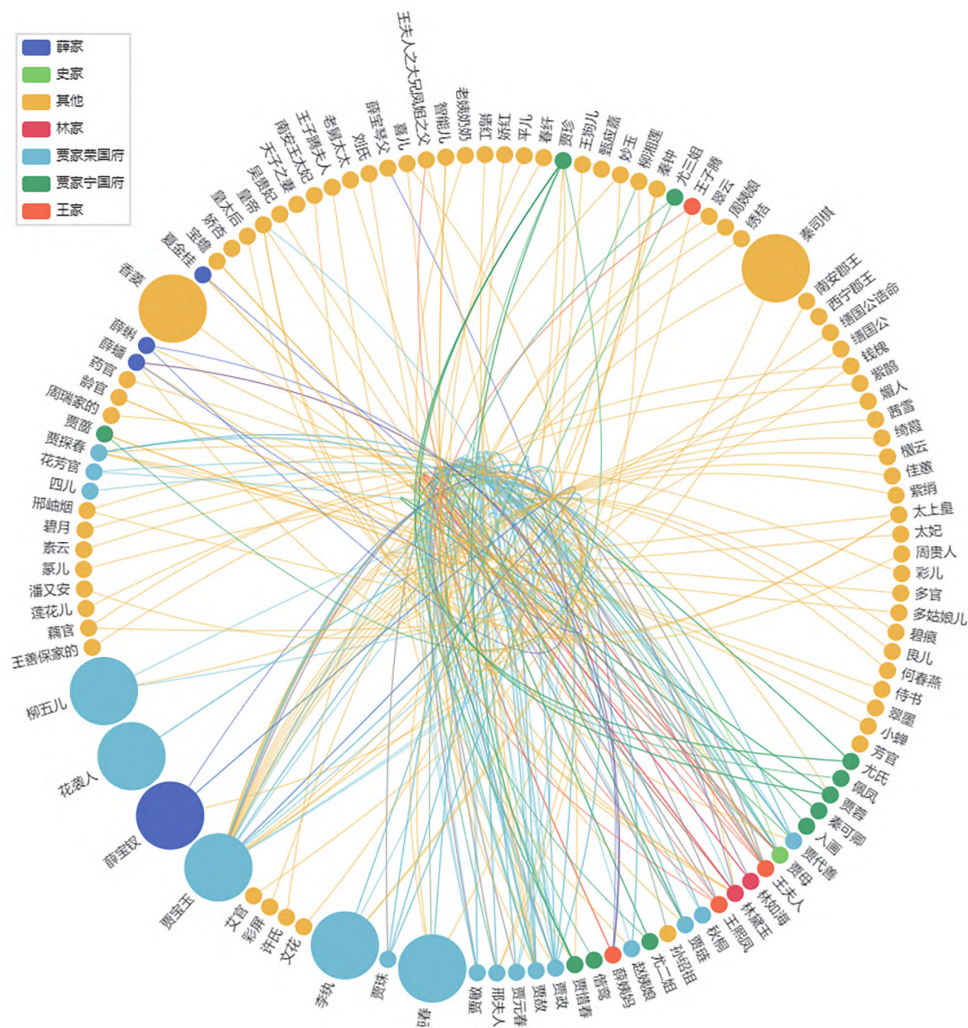


图 4 基于传统贝叶斯模型的概率分布

图 5 为基于关系的贝叶斯模型结果。在关系加权贝叶斯模型中,概率最大的 8 位,依大小排序分别为秦司棋、花袭人、李纨、薛宝钗、贾迎春、柳五儿、贾宝玉和香菱。说明经过加权更新的后验概率不仅同样找出了满足相应证据的个体,而且能根据其关系网络的一致性与其相关个体的后验概率对这些个体进行进一步的重要性排序。

表 2 为两种模型的概率估计排名。从推断准确性方面来看,在两种模型估计结果中,可能性最大的嫌疑人均均为秦司棋。但传统贝叶斯模型对最大嫌疑人的估计具有模糊性,无法准确地推断结果,而基于关系的贝叶斯模型对结果的推断更具准确性。从关系耦合方面来看,基于关系的贝叶斯模型能够

将秦司棋的相好潘又安的估计概率提高，而潘又安并不在大观园中，说明在缺乏其他证据时，仍能通过关键人物锁定相关的实体。

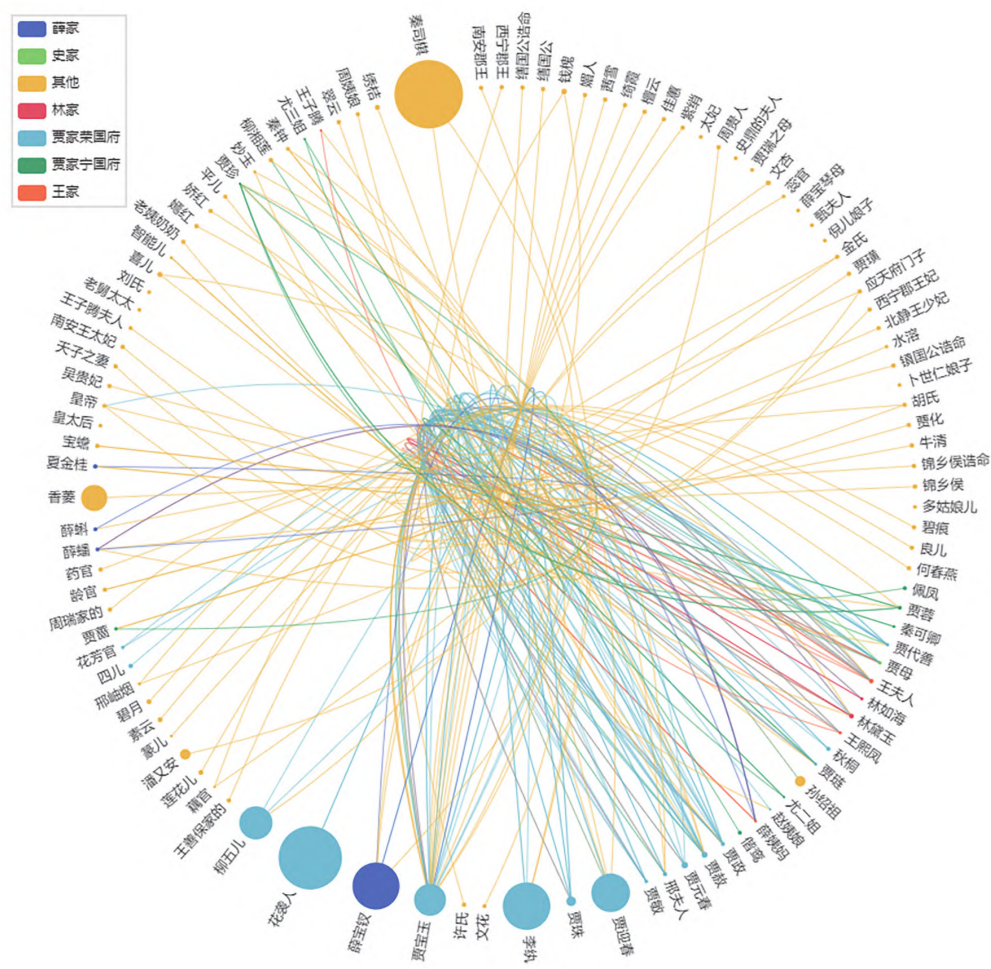


图 5 基于关系的贝叶斯模型下的概率分布

表 2 两种模型的后验概率

序号	贝叶斯模型		基于关系的贝叶斯模型		序号	贝叶斯模型		基于关系的贝叶斯模型	
	名字	概率/%	名字	概率/%		名字	概率/%	名字	概率/%
1	贾迎春	1.77	秦司棋	8.01	9	彩儿	0.44	孙绍祖	1.21
2	李纨	1.77	花袭人	7.42	10	多官	0.44	潘又安	1.21
3	贾宝玉	1.77	李纨	5.56	11	多姑娘	0.44	贾珠	1.11
4	薛宝钗	1.77	薛宝钗	5.50	14	何春燕	0.44	钱槐	0.61
5	花袭人	1.77	贾迎春	4.51
6	柳五儿	1.77	柳五儿	3.88	51	潘又安	0.44	良儿	0.50
7	香菱	1.77	贾宝玉	3.72
8	秦司棋	1.77	香菱	3.05	100	钱槐	0.44	贾琏	0.34

图 6 为基于关系的贝叶斯模型的关系网络图，节点大小为假设发生的概率，结构上分为两部分，一部分是以花袭人为概率中心的邻域网络，其邻域中较大概率的个体有薛宝钗、贾宝玉、李纨、柳五儿等，邻域中的个体均满足已有证据，且他们的关系为妻妾与丈夫、丫头与主人等亲密的关系；另一

部分是以秦司棋为概率中心的邻域网络,其邻域中较大概率率的节点有贾迎春、潘又安.而潘又安是秦司棋的相好,贾迎春是秦司棋的主人,均属于密切的关系.可见结合关系权重后的贝叶斯后验概率分布具有比较明显的“集群”现象,能锁定满足相关证据条件的部分个体,并根据网络的耦合关系对这些个体作进一步的计算,最终得到不同个体的嫌疑程度.

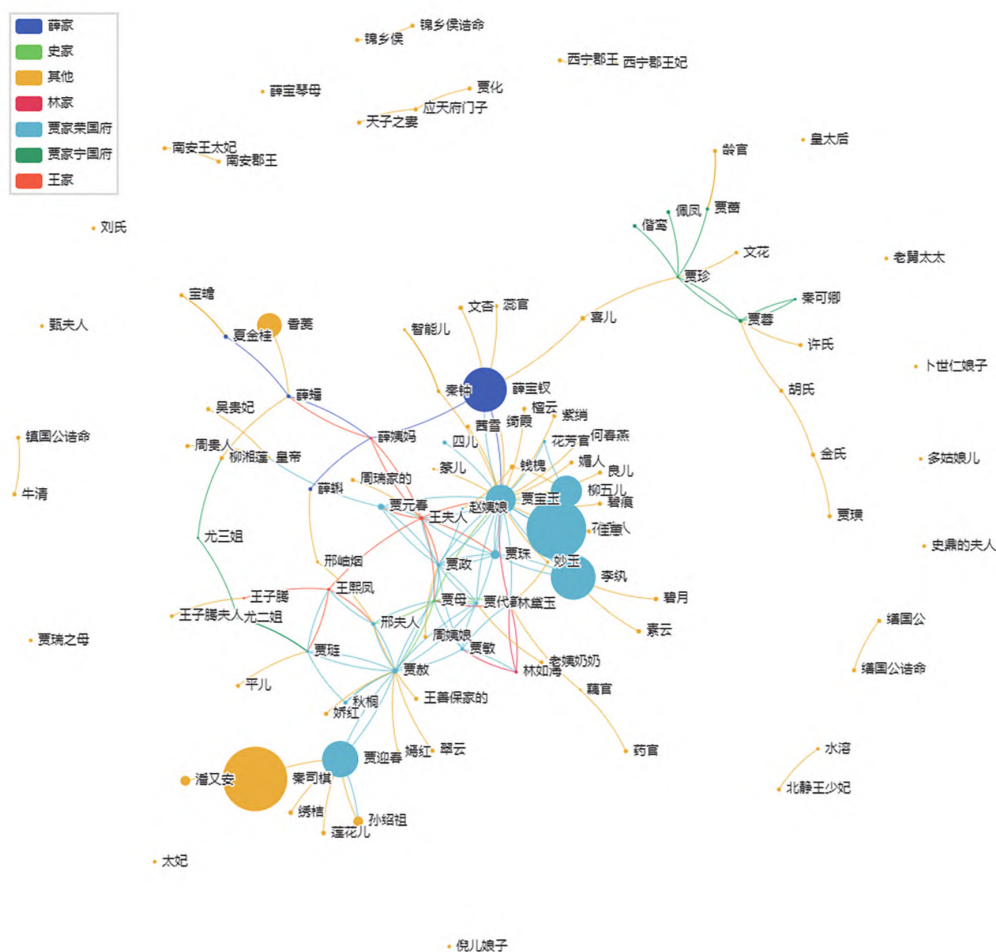


图 6 基于关系的贝叶斯模型下的图谱

3 结论与讨论

本文针对传统贝叶斯模型中参数之间可能存在相关关系的问题,提出了基于知识图谱的关系加权贝叶斯模型. 根据知识图谱节点属性自适应计算条件概率的方法,可以自动将定性属性转化为定量数据,提高了贝叶斯机制多次更新的速度,从而提高情报分析的效率;利用个体所在的社交关系网络,计算其所在网络的亲密关系程度,因此图谱的每个关系子网络均可自适应地转化为该网络的综合亲密权重. 该方法有效融合图谱中多种关系结构,减小参数之间的相关性对概率估计产生的偏差;在条件概率与亲密关系权重的基础上,完成了关系网络中的自适应加权贝叶斯推断算法,并推导了保证算法合理性的相关理论性质,实证结果表明该算法能较好地融入图谱的关系,给出更为准确的推断.

本文的亲密关系权重反映了社会关系一致性,在概率上表现为同增同减,在关系上表现为关系亲疏,如团伙犯罪中的同辈亲属关系应该设定较大权重。但是在复杂的社会关系网络中,模型还可以根据实体与其社会联系的互补性或竞争关系构建,例如家族继承中的利益相关关系应该设定较大权重,其嵌入的特征能进一步扩充,推断的可解释性进一步提升。探讨社会联系是刑侦推断与决策的重要课题,本文提出知识图谱嵌入贝叶斯模型的一般算法,操作简单,稳定性强,准确性高。其中,知识图谱

提供了强大的属性特征与关系特征,为公安情报分析提供高效、科学的支撑,是一个较好的切入点,有一定的推广价值。

参考文献

- [1]周顺忠,薛宏伟.科技兴警,助推公安工作质量与效率变革[N].光明日报,2021-03-29(02).
- [2]Lindley D V. Fiducial distributions and Bayes' theorem[J]. Journal of the Royal Statistical Society, Series B: Methodological, 1958, 20(1): 102-107.
- [3]Aljumily R. Quantitative criminology: Bayesian statistics for measuring the "Dark Figure" of crime[Z/OL]. [2024-03-20]. <https://ssrn.com/abstract=2999280>.
- [4]Pete Blair J, Rossmo D K. Evidence in context: Bayes' theorem and investigations[J]. Police Quarterly, 2010, 13(2): 123-135.
- [5]周长银,赵玉堂,孙亚星.基于贝叶斯信息更新的失事飞机发现概率模型[J].数学建模及其应用,2015,4(2): 71-78.
- [6]Valdés R M A, Comendador V F G, Sanz L P, et al. Prediction of aircraft safety incidents using Bayesian inference and hierarchical structures[J]. Safety Science, 2018, 104: 216-230.
- [7]魏新蕾,顾金尧,石拓,等.基于转移概率矩阵自学习的犯罪分布预测[J].北京理工大学学报,2020,40(1): 98-104.
- [8]Salo B, Sirén J, Corander J, et al. Using Bayes' theorem in behavioural crime linking of serial homicide[J]. Legal and Criminological Psychology, 2013, 18(2): 356-370.
- [9]Jonakait R N. When blood is their argument: probabilities in criminal cases, genetic markers, and, once again, Bayes' theorem[J]. University of Illinois Law Review, 1983: 369-422.
- [10]Anwar S, Loughran T A. Testing a Bayesian learning theory of deterrence among serious juvenile offenders[J]. Criminology, 2011, 49(3): 667-698.
- [11]闫召华.概率原理在犯罪嫌疑人摸排中的应用[J].中国人民公安大学学报:社会科学版,2013,29(5): 64-71.
- [12]陈雨楠,陈刚.基于贝叶斯理论的公安情报预测方法[J].齐鲁工业大学学报,2019,33(2): 69-77.
- [13]陈亮,张杨,黄一宁,等.公安情报的贝叶斯推理机制研究[J].山西大学学报:自然科学版,2010,33(1): 62-66.
- [14]黄忠良,翁文国.典型拐卖妇女犯罪团伙网络特征分析[J].清华大学学报:自然科学版,2020,60(4): 328-333.
- [15]骆公志,李震,黄卫东.加权先验概率优势关系的粗糙决策分析模型[J].统计与决策,2015,440(20): 67-70.
- [16]江旭晖,沈英汉,李子健,等.社交知识图谱研究综述[J].计算机学报,2023,46(2): 304-330.
- [17]Zou X. A survey on application of knowledge graph[C]//Proceedings of 4th International Conference on Control Engineering and Artificial Intelligence (CCEAI 2020), 2020, 1487(1): 012016.
- [18]周鹏,胡运涛,苗世茂.知识图谱在公安领域的构建与应用研究[C]//中国指挥与控制学会(Chinese Institute of Command and Control).第八届中国指挥控制大会论文集.北京:兵器工业出版社,2020: 386-392.
- [19]樊舒.支持智能应用的公安情报知识图谱本体构建[J].中国人民公安大学学报:自然科学版,2021,27(1): 28-34.
- [20]王蒙.《搜检大观园》评说[J].文学遗产,1990,2: 109-119.
- [21]俞晓红.从“抄检大观园”说《红楼梦》整体阅读观——《红楼梦》整本书阅读之五[J].学语文,2020,224(5): 6-10.
- [22]刘荣林.借勺水兴洪波误拾绣春囊,惑奸谗大抄检反现眼烧身——《红楼梦》小说创作美学系列论文“红楼艺鉴”之三[J].名作欣赏,2008,250(10): 127-130.
- [23]何先慧.只因一着错,满盘皆是输——论司棋与潘又安爱情悲剧的原因[C]//贵州省社会科学界联合会.红楼纵横谈——2012年贵州省社会科学学术年会第二十二分会场暨贵州省《红楼梦》研究学会2012年学术年会论文集.贵阳:贵州教育出版社,2012: 116-124.

Criminal Investigation Inference Model Based on Knowledge Graph and Weighted Bayesian Mechanism

LI Bo¹, ZHANG Yong^{1,2}, HU Yuqian¹

(1. School of Mathematics and Statistics, Central China Normal University, Wuhan, Hubei 430079, China;

2. Hefei NO.6 High School, Hefei, Anhui 230061, China)

Abstract: With the rapid development of informatization and digitization, various artificial intelligence methods have been extensively applied in the field of modern public security. Among them, knowledge graphs, with their powerful knowledge representation and mathematical modeling capabilities, have greatly improved the interpretability of intelligent decision-making, making them a hot topic of research and application in this field. Bayesian inference models based on the Bayesian mechanism have always been important means for public security intelligence and criminal investigation analysis. In this paper, we introduce the knowledge graph of public security into the traditional Bayesian model. By fully considering the coupling of parameters, we effectively update the posterior probability with weighted updates, significantly improving the accuracy of inference. Taking the case of "Embroidered Spring Bag" in the novel "The Story of the Stone" as an empirical analysis, this paper scientifically reveals the truth of a classic case. The proposed scheme of using knowledge graph's search adaptive weighting effectively improves the efficiency of the Bayesian model, making it a beneficial attempt to empower intelligent public security with artificial intelligence.

Key words: knowledge graph; Bayesian model; relationship weights; adaptive weighting

作者简介

李 波(1977—), 男, 博士, 教授, 研究方向为教育大数据、应用统计等。

章 勇(1989—), 男, 华中师范大学数学与统计学学院博士研究生, 合肥六中教师, 研究方向为教育大数据。

胡誉骞(1998—), 男, 硕士在读, 研究方向为教育大数据。