

法律事实认定的贝叶斯人工智能模型

刘 海

摘要:自由心证面临难以厘清证据、隐性主观判断和内心确信模糊三重困境。贝叶斯人工智能模型作为一种可视化、透明化和数量化的智能工具,可以辅助事实认定者完成心证过程,揭开心证黑箱,提升事实认定的准确性。它以贝叶斯网络图为知识系统,可视化呈现证据与事实之间的推论关系;以概率数字为数据系统,透明化展示证据与事实之间的依赖程度;以概率算法为推理系统,智能化计算出后验概率,数量化测量内心确信的推论强度。然而,它也存在局限性。一方面,它需要大量概率数字作为输入,但事实认定者对证据的主观判断却较难精准地转化为概率数字,从而影响模型输出结果的准确性。另一方面,它是显性表达主观性的工具,但不能消除主观性。

关键词:自由心证; 贝叶斯人工智能; 事实认定

DOI: 10.19836/j.cnki.37-1100/c.2024.03.016

事实认定包含举证、质证和认证三个过程,它是法律适用的基础,对作出公正的司法裁决起到关键作用。认证是指事实认定者的认知过程,是对事实认定的证成。在认证过程中,事实认定者对已采纳的证据进行推理和评价,从而决定相信什么,认定什么主张成立^①。评价证据和推断事实主要由事实认定者自由心证完成。但自由心证面临难以厘清证据、隐性主观判断和内心确信模糊三重困境。第一,大量证据与要件事实之间的错综复杂关系让事实认定者内心难以准确厘清证据和发现真相。第二,事实认定者在评价证据和推断事实的过程中,不可避免地涉及直觉、经验和个体认知等主观判断。这些“内隐的”主观因素阻碍准确地认定事实真相。第三,基于证据,事实认定者相信案件事实真相的内心确信程度具有模糊性。这种内心确信的强度被称为推论强度。是否可以数量化表示推论强度,并进行智能化计算?在司法实践中,这个问题是事实认定者最关心的问题,也是最重要的问题。

贝叶斯人工智能模型为回答这个问题提供了框架。它受知识和数据双驱动,以知识、数据、算法和算力四个要素来构造智能,属于第三代人工智能^②,具有更强的稳健性和可解释性。首先,它以贝叶斯网络图为载体建构知识系统。事实认定者结合案情,依据因果关系或证据关系,编织所有证据和待证事实之间的结构,搭建贝叶斯网络图。贝叶斯网络图以可视化方式展示了大量证据与事实之间的结构联系。其次,它以概率数字建构数据系统。事实认定者发挥主观能动性,根据经验法则、科学法则和因果联系,评估证据和待证事实之间依赖关系的强弱,创建概率分布表。概率数字“透明地”呈现了事实认定者对证据的主观判断。最后,它以概率算法建构推理系统。概率算法使概率计算自动化和智能化,自动更新概率数值,得到想要的后验概率。后验概率数量化表示了事实认定者对案件真相的内心确信程度。

基金项目:国家社科基金一般项目“司法证明的逻辑研究”(23BZX127);2022年教育部高等学校一般国内访问学者项目“贝叶斯概率在建构证据评价智能化模型中的应用”。

作者简介:刘海,上海对外经贸大学统计与信息学院副教授,硕士生导师(上海 201620; liuhai@suibe.edu.cn)。

① 何福来:《证据法哲学——在探究真相的过程中实现正义》,樊传明、曹佳、张保生等译,北京:中国人民大学出版社,2021年,第45页。

② 张钹、朱军、苏航:《迈向第三代人工智能》,《中国科学:信息科学》2020年第9期。

一、证据与事实之间推论关系的贝叶斯网络图

事实认定中的证据推理是一种常识推理,涉及知识表示和知识推理。贝叶斯人工智能以贝叶斯网络图结构化表示知识,将大量证据和事实之间的联系可视化。贝叶斯网络图由节点和弧构成,如图1所示。节点代表变量。有些变量表示不可观察的未知假设,即待证事实,如“张某有罪(H1)”。有些变量表示可以观察的事实,即证据,如“血型DNA被检测为匹配(E1)”,这个证据可以通过法医鉴定技术来观察。弧是一种有方向的边,由一个节点指向另一个节点,它表示变量之间的“直接”依赖关系,这种关系通常是因果关系,也可以是概率影响等相关关系^①。因此,贝叶斯网络图是表示和沟通多个不同假设和证据之间关系的自然方式。如果从节点A有一条弧指向节点B,则称节点A是节点B的父节点,B为A的子节点。在贝叶斯网络图中,如果两个节点之间没有弧直接相连,那么它们存在某种条件独立性。条件独立性极大地降低概率计算的复杂度,是贝叶斯网络模型的优势之一。

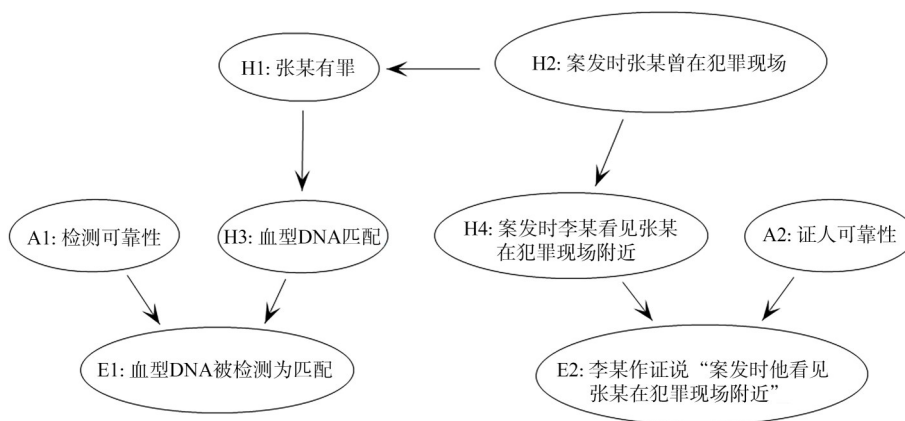


图1 DNA鉴定意见和证人证言的贝叶斯网络图^②

(资料来源:作者自制。)

2013年,芬顿等人通过创建一系列被称为“习语”的推理模式,开创了一种基于习语的方法,运用它可以轻松地法律案件建构一个“共识的”贝叶斯网络图^③。习语是一种特定的贝叶斯网络图片段,它代表法律不确定推理中人们经常使用的推理类型。这些习语可以重复使用,从而加快搭建贝叶斯网络图的过程,并获得更高质量的图结构。这种基于习语的方法已被应用于研究实际法律案件,呈现和分析整个案件的证据推理过程,辅助事实认定者认定案件事实^④。这些实际案例研究表明,基于习语的方法可以有效地建立共识的贝叶斯网络图,从而具有实用性。

在法律证据推理中,存在6种常见的习语:证据习语、证据依赖习语、证据准确性习语、动机习语、机会习语和不在场证据习语。运用这些习语,事实认定者可以快速地创建整个案件的贝叶斯网络图。在认证过程中,事实认定者在考虑控辩双方提出的证据、事实主张和论证后,利用贝叶斯网络图结构

^① Fenton N., Neil M., *Risk Assessment and Decision Analysis with Bayesian Networks*, London: CRC Press, 2019, p. 165.

^② 图1中的节点分为三类:H表示假设(Hypothesis)节点,E表示证据(Evidence)节点,A表示证据的可靠性或准确性(Accuracy)节点。数字1、2、3、4表示每类节点的序号。

^③ Fenton N., Neil M., Lagnado D. A., “A General Structure for Legal Arguments About Evidence Using Bayesian Networks”, *Cognitive Science*, 2013, 37(1), pp. 67-102.

^④ Vlek C. S., Prakken H., Renooij S., et al., “A Method for Explaining Bayesian Networks for Legal Evidence with Scenarios”, *Artificial Intelligence and Law*, 2016, 24(3), pp. 285-324.

化表示大量证据和待证事实之间的联系,使其心证公开化、可视化。这样有助于展现事实认定者在心证时可能犯的推理错误,从而提升事实认定的准确性,实现公平和正义。

针对刑事案件,运用习语建构贝叶斯网络图分为三个步骤。为了清晰地阐释这三个步骤,下文将以评价DNA鉴定意见和证人证言为例来说明其运作原理。假设发生了一起谋杀案,犯罪嫌疑人张某被捕。现假定只有两项证据,一是DNA鉴定意见,法医鉴定意见指出:张某血型DNA与犯罪现场收集到的血型DNA相匹配;二是证人证言,证人李某作证说道:案发时他看见张某在犯罪现场附近。那么,如何可视化展示这两项证据与最终待证事实之间的结构联系呢?

首先,事实认定者结合案情,识别最终待证事实(即控方假设),以及犯罪动机和犯罪机会。在上述案例中,最终待证事实为“张某有罪”,即张某实施了谋杀行为,在图1中用节点H1表示。在贝叶斯网络图中,每个节点代表一个变量,变量有取值或状态。由于证据推理是基于有限知识的推理,因此每个变量的状态具有不确定性,概率测度了变量处于某个状态时的不确定程度。根据状态类型,变量分为离散变量和连续变量,离散变量又细分为布尔变量、整数值变量和多项类别变量。图1中的每个节点都是布尔变量,它们有两种状态:为真或为假。H1为真代表“张某有罪”成立,为假则表示“张某有罪”不成立。H1到底“为真”还是“为假”,具有不确定性。节点H2是机会变量,表示“案发时张某曾在犯罪现场”。本案例中不存在动机证据,因此没有动机节点。

其次,事实认定者要辨识案件中已采纳的证据,以及最终待证事实相关的其他假设。对于最终待证事实H1,本案例不存在与之相关联的直接证据,但有一个与它相关的假设H3作为子节点,和机会节点H2作为其父节点。H2指向H1,表达的因果关系是:由于案发时张某在犯罪现场,因此张某有可能实施谋杀行为。设E1表示鉴定意见“血型DNA被检测为匹配”,即经法医专家鉴定,张某血型DNA与犯罪现场收集到的血型DNA相匹配。E2表示证人证言“李某作证说案发时他看见张某在犯罪现场附近”。E1是H1的间接证据,E2是机会节点H2的间接证据。证据E1经过假设H3与H1相关联,H3表示鉴定意见的结论:“血型DNA匹配”,即张某血型DNA与犯罪现场收集到的血型DNA相匹配。证据E2经由假设H4与H2相关联,H4表示证人证言的内容:“案发时李某看见张某在犯罪现场附近”。推论链“H1→H3→E1”表达的因果关系是:由于张某实施了谋杀,他很可能在现场留下血迹,因此张某的血型DNA会与犯罪现场收集到的血型DNA相匹配,进而法医专家鉴定出两个血型DNA相匹配。推论链“H2→H4→E2”呈现的因果关系是:由于案发时张某在犯罪现场,因此李某看见张某在犯罪现场附近,于是张某作证说他看到张某在犯罪现场附近。

在分析证据时,有必要区分证据本身和证据所断言的事实,这种事实被称为“证据性事实”。鉴定意见是证据,鉴定意见表达的结论是证据性事实。类似地,证人证言是证据,证言所描述的内容是证据性事实。证据性事实不可观察,是一种假设,具有不确定性,需由证据来证明。然而,不同的证据往往存在不同的可靠程度。可靠性影响证据的推论强度。设节点A1表示“检测可靠性”,A2表示“证人可靠性”,那么它们分别影响着证据E1和E2的推论强度,分别是证据E1和E2的父节点。

最后,事实认定者根据节点间的因果关系或相关关系,运用各种习语,建立证据和假设之间的结构联系,并在贝叶斯人工智能软件GeNIe中绘制完整的图形^①。图1清晰地呈现了事实认定者的心证推理路径,可视化展示了证据和假设之间的结构联系。

需要说明的是,本文主要目的是阐释贝叶斯人工智能应用于事实认定的运作机制,为了模型简洁并根据本案例的情况,图1中只用到了三种习语:证据习语(如H2和H4之间的结构),证据准确性习语(如A1、H3和E1之间的结构)和机会习语(H2和H1之间的结构)。由于证据之间相互独立,因而不涉及证据依赖习语。另外,本案例没有不在场证据和动机证据,因此也没有涉及动机习语和不在场证据习语。

① 这款软件由美国BayesFusion有限公司开发,稳定性高、可靠性强,广受业界认可,对学术研究者免费开放,但商业用途要收费,可在<https://www.bayesfusion.com/>下载,访问日期:2024年3月2日。

与其他任何模型类似,贝叶斯网络图也是对现实的近似描述。在选择节点数量和考虑两个节点之间是否应该画弧时,事实认定者需要在现实和效率之间做必要的权衡。节点和弧越多,越接近现实,但模型的创建和运行也越困难。尽管基于习语的方法提供了一套建构贝叶斯网络图的系统化手段,但证据、假设和习语的选择仍然涉及事实认定者的认知和主观判断,具有一定的主观性。为了建立一个能够达成共识的贝叶斯网络图,事实认定者可以与专家或同行交流讨论,不断修正和改进图的结构,最终促成共识。

二、证据与事实之间依赖程度的概率模型

贝叶斯网络图揭示的仅仅是证据和假设之间的推论关系。然而,推论强度如何?也就是说,基于全案证据,事实认定者可以在何种程度上相信最终待证事实的真相?回答这个问题还得求助于贝叶斯人工智能的数据系统。

在贝叶斯网络图中,每个节点都对应着一个概率分布表。概率分布表是由节点的每个状态和状态对应的概率数值构成的表格。对于无父节点的节点,如H2、A1和A2,它们的概率分布表由先验概率构成。对于有父节点的节点,其概率分布表则由条件先验概率构成。事实认定者基于案件的背景信息和经验法则,来评估先验概率和条件先验概率数值。先验概率透明化呈现事实认定者对假设所处状态(为真或为假)可能性大小的主观评估,条件先验概率则透明化展示事实认定者对父节点与子节点之间依赖关系强弱的主观评价。因此,概率分布表以数字形式透明地呈现了事实认定者对证据与事实的主观判断和信念。

先验概率和条件先验概率共同构成贝叶斯人工智能的数据系统,它们是贝叶斯人工智能的重要“输入”。另一个重要“输入”是证据的状态。基于数据系统,输入证据的状态,运用贝叶斯人工智能软件,就可以自动计算出后验概率。后验概率是推论强度的量化模型,其取值反映了推论强度的大小,是贝叶斯人工智能最重要的“输出”。后验概率越大,推论强度就越高,基于证据事实认定者相信最终待证事实成立的程度就越强。反之,则越弱。

(一) 概率评估的困难与应对

创建数据系统就是给贝叶斯网络图中每个节点确定对应的概率分布表,这涉及先验概率和条件先验概率数字的来源和计算方式。在大多数应用领域中,可以从各种来源获得概率信息。最常见的三种来源是统计数据、文献资料和人类专家^①。在事实认定领域中,除了DNA等少量科学证据有统计数据库外,其他绝大部分证据都无法找到数据信息,也缺乏包含相关概率数字的文献资料。因此,概率分布表主要由事实认定者依据自己或他人的专业知识和经验来主观估算。事实认定者也可以与其他相关专家或同行交流、讨论,最后形成能达成共识的概率评估。专业知识和经验不仅可以帮助评估所需的概率,还可以对从其他来源获得的概率进行微调,让它适用于当前案件的具体情况,并在贝叶斯网络中验证数字。基于人类的专业知识和经验引出概率数字是一个热门研究话题,其过程很困难,但不是不可行。虽然有学者批判道:主观概率“可能是任何数字,而且它们不需要以任何方式受到证据质量的限制”^②。然而,这个批判有失偏颇。实际上,贝叶斯网络图本身对合理假设概率的可行范围设置了许多较强约束,因为它编码了变量间的部分或全部因果关系,这种关系使得比单独引出概

① Druzdzel M. J., van der Gaag L. C., “Building Probabilistic Networks: ‘Where Do the Numbers Come From?’”, *IEEE Transactions on Knowledge and Data Engineering*, 2000, 12(4), pp. 481-486.

② Allen R. J., Pardo M. S., “Relative Plausibility and Its Critics”, *The International Journal of Evidence & Proof*, 2019, 23, pp. 5-59.

率更容易和更准确^①。“概率数字的主观性是一种被夸大的障碍。”^②在任何案件中,无论采用什么方法分析证据,事实认定者都会涉及个人主观判断。贝叶斯人工智能方法只是简单地用一种公开接受测试和挑战的数字形式透明化表达了内含的隐性主观性。

运用贝叶斯人工智能模型也可能面临这样一种风险:为了得到某个想要的概率输出结果,事实认定者“反推式”地设定先验概率,从而让其心证过程披上“逻辑”的外衣,让人们误以为其推理过程非常理性。然而,这种风险不是贝叶斯人工智能模型的缺陷,而是事实认定者不当运用造成的后果。贝叶斯人工智能模型是一种结构化的工具,它可以让推理过程遵循逻辑,不犯逻辑错误,还能辅助事实认定者可视化、透明化、数量化展示其心证推理过程,这种显性的展示有助于他人和事实认定者自己清晰地看到心证过程,从而就心证展开讨论和交流,评价心证过程是否合理、该如何改进、是否能达成一致和获得共识。在这种意义上,贝叶斯人工智能模型确实可以提升事实认定的准确性。但是,它不能决定事实认定者的心证,事实认定者在心证过程中始终占主导地位。如果事实认定者对证据的评价不够准确,甚至故意曲解证据,然后再用贝叶斯人工智能模型展示这种有瑕疵的心证过程,那么所得出的结果也必定是不准确的。心证推理过程自身的准确性决定了贝叶斯人工智能模型输出结果的准确性,模型只是帮助事实认定者显性地展现心证而已。

(二)DNA 鉴定意见和证人证言的概率评估

根据概率的计算方法,我们创建了如表1所示的8个概率分布表,分别对应着图1中8个节点。表1总共包含34个概率数字,其中17个概率是真正要评估和确定的数字,依据概率互补公理,可计算得到另外17个概率。“T”和“F”分别表示“为真”和“为假”两种状态。根据贝叶斯定理,这34个概率数字就是图1案例所需要的全部概率输入,有了这些输入就可以进行概率推理,计算想要的目标概率输出。

表1 DNA 鉴定意见和证人证言的数据系统

H2=T		0.1	A1=T		0.9	A2=T		0.7
H2=F		0.9	A1=F		0.1	A2=F		0.3

H2	T	F	H1	T	F	H2	T	F
H1=T	0.1	0	H3=T	1	0.0003	H4=T	1	0
H1=F	0.9	1	H3=F	0	0.9997	H4=F	0	1

A1	T		F		A2	T		F	
H3	T	F	T	F	H4	T	F	T	F
E1=T	1	0	0.98	0.03	E2=T	1	0	0.5	0.5
E1=F	0	1	0.02	0.97	E2=F	0	1	0.5	0.5

资料来源:作者自制。

表1中第1个表格是机会节点H2的先验概率分布表,第4个表格是最终待证事实H1的条件先验概率分布表。芬顿等人通过定义“犯罪现场”“犯罪时间”“扩展的犯罪现场”和“扩展的犯罪时间”等概念,提供了在犯罪现场和犯罪时间内“有作案机会的总人数 n ”,以及在扩展的犯罪现场和犯罪时间内“与被告人具有相同或相似作案机会的总人数 N ”的方法。根据这两个数据,芬顿等人认为,“被告人

① Fenton N., Neil M., *Risk Assessment and Decision Analysis with Bayesian Networks*, London: CRC Press, 2019, p. 292.

② Fenton N., Neil M., Yet B., et al., “Analyzing the Simonshaven Case Using Bayesian Networks”, *Topics in Cognitive Science*, 2020, 12(4), pp. 1092-1114.

在犯罪现场”的先验概率是 n/N ;在被告人有犯罪机会的条件下,“被告人是凶手”的条件先验概率为 $1/n$;于是“被告人是凶手”的先验概率为 $1/N$ ^①。将此方法应用于图1案例中,事实认定者可以依据警方或法医提供的数据信息来评估 H_2 和 H_1 的概率分布表。

首先,法医对死者进行尸检分析,推测具体的犯罪时间。警方侦查犯罪现场,推测在犯罪时间内有多少人到过犯罪现场。这样就确定了有作案机会的总人数 n ,比如 $n=10$ 。其次,由于在这10个人中,警方往往不能确定嫌犯张某是其中的某个人,于是警方需要扩大搜索的空间和时间范围,确定扩展的犯罪现场和犯罪时间。扩展的犯罪现场和犯罪时间必须满足两个条件。第一,在这个范围内搜索确定的总人数 N 中,必须包含被告人张某。第二,这个搜索范围是包含张某在内的距离犯罪现场最近和犯罪时间最短的范围。假定在本案例中, $N=100$ 。也就是说,在扩展的犯罪现场和犯罪时间内,警方确定了100个有机会实施谋杀的人,并且犯罪嫌疑人张某就是其中之一。最后,根据芬顿等人给出的计算公式,可以得到第1个表格和第4个表格中的概率数值。

在第1个表格中, $P(H_2=T)=10/100=0.1$,即根据上述数据,“张某曾在犯罪现场”的先验概率为0.1。依据概率互补公理, $P(H_2=F)=1-0.1=0.9$,即“张某不在犯罪现场”的概率为0.9。

在第4个表格第2列中, $P(H_1=T|H_2=T)=1/10=0.1$,即在“张某曾在犯罪现场”的条件下,“张某有罪”的概率是0.1。根据概率互补公理, $P(H_1=F|H_2=T)=1-0.1=0.9$ 。依据变量间的因果关系,第3列表示在“ H_2 为假”,即张某不在犯罪现场的情况下,“ H_1 为真”的概率是0,“ H_1 为假”的概率是1。

第2个表格是检测可靠性节点 A_1 的先验概率分布表,它表示法医鉴定意见可靠的概率是0.9,不可靠的概率是0.1。鉴定意见可靠性的概率可以通过统计法医专家过往检测准确性的历史频率数据来估算。如果历史数据显示这名法医专家的鉴定准确率是0.9,那么事实认定者应该90%地确信鉴定意见是准确的。

第3个表格是证人可靠性节点 A_2 的先验概率分布表,它表示事实认定者70%相信证人说真话,30%认为证人说假话。根据证人在法庭作证时的行为表现、关于证人的其他品格信息,以及自己的经验,事实认定者可以主观评估证人可靠性的概率。

第5个表格是血型DNA匹配节点 H_3 的条件先验概率分布表。根据因果关系,当张某有罪时,张某的血型DNA非常可能与现场收集到血型DNA相匹配,于是得到第2列的概率数值1和0。当张某不是凶手时,张某的血型DNA与现场收集到血型DNA相匹配的概率可以通过“随机匹配概率”来估计。如果案件发生在甲市,且嫌犯张某是甲市的男性常住人口,那么可以在DNA数据库中,调取包含张某在内的甲市所有男性的血型DNA,比如共有 x 个DNA样本。然后分析犯罪现场收集到的血型DNA,识别出它各个特征点位的基因型,并统计每个基因型在这 x 个DNA样本库中出现的频率。由于DNA每个特征点位的基因型是独立的,因此将这些频率相乘,所得结果就是随机匹配概率。缘于DNA的特异性,随机匹配概率通常很小。在本案例中,我们假设随机匹配概率为0.0003,即在“ H_1 为假”的条件下,“ H_3 为真”的概率是0.0003。根据概率互补公理,得到 $P(H_3=F|H_1=F)=0.9997$ 。

第6个表格是节点 H_4 的条件先验概率分布表。当张某曾在犯罪现场时,证人李某很可能看到张某在犯罪现场附近,根据因果关系,得到第2列的概率数值1和0。类似地,当张某不曾在犯罪现场时,证人李某不可能看到张某,由此可得到第3列的概率数值0和1。

第7个表格是证据节点 E_1 的条件先验概率分布表。根据因果关系,在“ A_1 为真且 H_3 为真”,即检测可靠且血型DNA相匹配的情况下,血型DNA必然被检测为相匹配,因此“ E_1 为真”的概率是1,“ E_1 为假”的概率是0。同理得到第3列的概率数值0和1。在“ A_1 为假”,即检测不可靠的情况下,法

① Fenton N., Lagnado D., Dahlman C., et al., “The Opportunity Prior: a Proof-Based Prior for Criminal Cases”, *Law, Probability and Risk*, 2019, 18(4), pp. 237-253.

医专家通常犯两类错误,一是错误否定,二是错误肯定。错误否定是将本来相匹配的DNA鉴定为不匹配,错误肯定是将本来不匹配的DNA鉴定为匹配。这两类错误的概率可以通过历史频率数据来估算。本案例假设了错误否定的概率为0.02,错误肯定的概率为0.03,再结合概率互补公理,就得到了第4列和第5列的概率数值。

第8个表格是证据节点E2的条件先验概率分布表。根据因果关系,在“A2为真且H2为真”,即证人李某可靠且案发时李某看到张某在犯罪现场附近的情况下,李某在法庭肯定会作证说他曾看到张某,因此“E2为真”的概率是1,“E2为假”的概率是0。类似可得第3列的概率数值0和1。在“A2为假”,即证人李某不可靠的情况下,证人是否说真话较难评估。在无任何信息情况下,根据最大熵原则^①,本案例假设概率是“等可能的”,由此得到第4列和第5列的概率数值0.5。

由此可见,表1中的34个概率数字都是建立在公开和透明的数据或假设之上获得的,它们透明地展现了事实认定者对案件的主观判断,因此贝叶斯人工智能具有“可解释性”。这有助于暴露事实认定过程中可能犯的误差,促成人们在审查、交流和讨论中修正自己的观点并最终达成共识,提升事实认定的准确性,并为进一步将知识和经验转化为计算奠定基础。另外,创建概率分布表只涉及对先验概率和条件先验概率等概念的理解,将知识和经验等主观判断转化为概率数字,并不需要复杂的概率计算,因此,即使事实认定者不太擅长概率推理,也能较好地建构概率分布表。

三、证据与事实之间推论强度的智能化计算

除了贝叶斯网络图和概率分布表清晰地展示大量证据和假设之间的依赖关系和依赖程度外,贝叶斯人工智能的最大优势还在于基于它们所进行的贝叶斯网络推理,概率算法使得推理过程自动化和智能化。贝叶斯人工智能提供了一套严格的系统方法来传播证据和计算证据对事实的推论强度,帮助事实认定者形成准确和翔实的意见,从而决定相信什么和认定什么。“恰当地使用贝叶斯网络推理能够使证据的相关性得到有意义的评估和交流,从而有可能大大提高刑事司法系统的效率、透明度和公正性以及裁决的准确性。”^②

按照推理的方向,贝叶斯网络推理分为两种类型。一是从原因到结果的推理,这种推理遵循网络图中弧的方向,也被称为预测推理。二是诊断推理,它是从结果、现象或证据到原因的推理,这类推理的方向与弧的方向相反。例如,在图1中,当观察到证据E1的状态为真时,我们对H1为真的信念将增加,即后验概率 $P(H1=T|E1=T)$ 将大于先验概率 $P(H1=T)$ 。在事实认定过程中,人们经常使用诊断推理,在观察到证据后修正对假设的信念。通常,律师、陪审员或法官从一些关于最终待证事实(如被告人有罪)的先验假设开始,如无罪推定,即假设“被告人并不比其他有行为能力的人更有可能有罪”,然后基于获得的证据更新对最终待证事实的信念。这个过程与贝叶斯定理高度契合,因此从这个意义上讲,贝叶斯人工智能方法应用于事实认定是自然的。

贝叶斯网络推理涉及大量的概率计算,对使用者的数学技能要求很高。即使只有少量节点的贝叶斯网络图,执行正确概率推理所需的计算也太过复杂,对于数学技能较高的使用者来说,他们也无法手工完成计算,更何况是对数学可能比较陌生的事实认定者。概率计算的复杂性是造成贝叶斯网络模型早期很少应用于法律审判实践的主要原因之一。然而,20世纪80年代末,算法的突破使得对绝大多数的贝叶斯网络都能自动地执行有效的概率推理。这些算法随后被纳入广泛可用的软件工具中,如GeNIe软件,用户无需掌握太多的数学和统计知识就可以创建和运行贝叶斯网络模型。概率计算完全由软件工具自动完成,无需手工操作。这就实现了概率推理的自动化与智能化。

① Jaynes E. T., “Information Theory and Statistical Mechanics”, *Physical Review*, 1957, 106(4), pp. 620-630.

② Fenton N., Neil M., *Risk Assessment and Decision Analysis with Bayesian Networks*, London: CRC Press, 2019, p. 493.

首先,贝叶斯网络推理使在连续推论中推论强度的传播和更新逻辑化、科学化。在表1所示的数据系统基础上,运用GeNIe软件运行图1所示的贝叶斯网络图,所得部分结果如表2所示。表2列举了DNA鉴定意见E1和证人证言E2对最受关注的两个假设“案发时张某曾在犯罪现场H2”和“张某有罪H1”的影响效果。表2第2行表示先验概率 $P(H2=T)=10\%$, $P(H1=T)=1\%$ 。也就是说,在不考虑证据E1和E2时,基于背景信息知识,事实认定者10%地确信案发时张某曾在犯罪现场,1%地确信张某有罪。第3行表示后验概率 $P(H2=T|E1=T)=77.5875\%$, $P(H1=T|E1=T)=75.3463\%$ 。这说明在只考虑证据E1时,事实认定者相信“H2为真”的确信度由原来的10%提高到了77.5875%,相信“H1为真”的确信度由原来的1%提高到了75.3463%。第4行表示后验概率 $P(H2=T|E2=T)=38.6364\%$, $P(H1=T|E2=T)=3.8636\%$ 。这表示在只考虑证据E2,事实认定者相信“H2为真”的确信度由原来的10%提高到了38.6364%,相信“H1为真”的确信度由原来的1%提高到了3.8636%。由此可见,对于假设H2和H1,证据E1的推论强度分别大于证据E2。第5行表示后验概率 $P(H2=T|E1=T,E2=T)=95.1496\%$, $P(H1=T|E1=T,E2=T)=92.4010\%$ 。这说明,在同时考虑两个相互补强证据E1和E2时,它们互相印证,推论强度比只考虑单个证据时都大。后验概率数值的变化清晰地反映证据相互补强的效果。在同时考虑证据E1和E2时,事实认定者95.1496%地确信案发时张某曾在犯罪现场,92.4010%地确信张某有罪,确信度非常高。

表2 DNA鉴定意见和证人证言在连续推论中推论强度的传播与更新

证据输入	“H2=T”的概率	“H1=T”的概率
无证据输入	10.0000%	1.0000%
输入E1=T	77.5875%	75.3463%
输入E2=T	38.6364%	3.8636%
输入E1=T,E2=T	95.1496%	92.4010%
输入E1=F,E2=T	36.1754%	0.0081%
输入E1=T,E2=F	37.9231%	36.8276%

资料来源:作者自制。

其次,贝叶斯网络推理不仅能处理补强证据,还能统一分析冲突证据。后验概率数值的变化以透明化方式彰显证据相互补强和相互冲突的效果。表2第6行表示后验概率 $P(H2=T|E1=F,E2=T)=36.1754\%$, $P(H1=T|E1=F,E2=T)=0.0081\%$ 。其中,“E1=F”表示被观察到的证据是与E1相反的证据,即“血型DNA被检测为不匹配”,“E2=T”表示证据E2本身,即“李某作证说案发时他看见张某在犯罪现场附近”,这两个证据相互冲突。但贝叶斯网络推理能综合协调分析这两个冲突证据的推论强度。在同时考虑“E1=F,E2=T”的情况下,“H2为真”的概率由原来的10%提高到了36.1754%,“H1为真”的概率由原来的1%降低到了0.0081%。然而,第7行显示 $P(H2=T|E1=T,E2=F)=37.9231\%$, $P(H1=T|E1=T,E2=F)=36.8276\%$ 。这说明在同时考虑“血型DNA被检测为匹配”和“李某作证说案发时他没有看见张某在犯罪现场附近”这两个冲突证据时,“H2为真”的概率由原来的10%提高到了37.9231%,”H1为真”的概率由原来的1%提高到了36.8276%。

表2中明确的数据结论是贝叶斯人工智能方法的最大优势。基于先验假设和观察到的证据,贝叶斯网络模型可以计算出明确的后验概率。它是法官或陪审员作出事实裁决的重要参考和依据。需要说明的是,这里的数据结论完全不同于“法定证据制度”模式中的数据要求。法定证据制度发源于中世纪时期的欧洲,法律预先规定了各种证据的证明力和判断证据的规则,诸如僧侣作证优于世俗者,两个典型的证人证言构成一个完全的证据等。事实认定者根据预先设定的规则,机械地计算证据数量的多少、证明力的强弱和推论强度的大小,并据此认定案件事实。法定证据制度已被现在盛行的

自由心证式证据自由评价制度所取代。在贝叶斯人工智能框架下,事实认定者对证据和待证事实自由评价,贝叶斯网络图和概率分布表都是事实认定者内心自由评价的可视化、数量化、透明化和公开化结果,这些结果不受法律规则的约束,而受事实认定者自身的知识和经验,以及逻辑规则的限制。因此,在这些结果的基础上,利用概率逻辑和概率算法,贝叶斯人工智能软件自动化和智能化计算出的后验概率也不具有法定性。贝叶斯人工智能模型的功效在于“辅助”自由心证,“揭开”心证黑箱,而非“排斥”心证过程。

四、结语和展望

本文阐述了贝叶斯人工智能应用于事实认定之认证阶段的运作原理,尽管是以刑事案件为例进行说明,但它也适用于民事案件,因为所有案件的证据推理逻辑是相同的,不同的只是证明标准上的差异。首先,建构知识系统。当所有证据经过举证和质证,并被采纳后,事实认定者根据证据与待证事实之间的因果或相关关系,运用基于习语的方法,在贝叶斯人工智能软件中绘制贝叶斯网络图,如“GeNIe”软件。其次,创建数据系统。基于统计数据、文献资料或者专业知识和经验这三个概率信息来源,事实认定者评估证据与待证事实之间依赖关系的强弱,给每个节点指派概率分布表。最后,输入证据状态,软件自动推理。“GeNIe”软件自带各种概率推理算法,输入证据状态后,它可以自动计算出待证事实的后验概率。后验概率是推论强度的计算模型,以数量化的形式回答事实认定者可以在“何种程度上”相信其主张的事实真相。在司法实践中,这个回答是事实认定者最想要的答案,也是最重要的答案,是做出事实裁决的参考和依据。

尽管贝叶斯人工智能目前仅在法庭科学领域获得了广泛应用,法医专家越来越多地将贝叶斯概率论作为理论框架,运用软件工具设计贝叶斯网络来分析证据^①。但我们相信,法律界未来终将敞开心扉,打破学科文化壁垒,接受贝叶斯分析框架,让它应用于更多的法律分支领域。主要存在五点理由:第一,贝叶斯人工智能不仅可以利用机器学习的方法处理大数据,它还可以整合专家知识和专家判断这样的智慧数据^②。在处理法庭审判这类数据资料较少或缺失的风险决策时,贝叶斯人工智能也能结合人类知识和经验提供可行的分析模型。第二,贝叶斯网络推理使在连续推论中证据之推论强度的传播和更新逻辑化、科学化。它不仅能处理补强证据,还能统一分析冲突证据。第三,软件工具的运用解决了事实认定者面临的“数学演算薄弱”之困境,助推了贝叶斯人工智能方法在事实认定实务中的应用。不太擅长概率推理的事实认定者,经过简单的培训也能很好地运用软件工具建构贝叶斯网络模型。第四,贝叶斯人工智能模型有助于打开心证黑箱,以可视化、透明化和数量化方式展现事实认定者的主观判断和推理过程,提升事实认定的准确性。第五,贝叶斯人工智能是人机协作的概率推理专家系统,其目标是“辅助人”和“增强人”,而不是“取代人”。所有司法决策仍是由人来完成,事实认定者的主体地位保持不变。贝叶斯人工智能的应用,使审判成为一种半自动化的人机协同操作,提高司法效率,缓解“案多人少”的司法资源紧张之现状,为审判体系和审判能力现代化注入新动力。

① Taroni F., Biedermann A., Bozza S., et al., *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*, UK: John Wiley & Sons, 2014, p. 85.

② Fenton N., Neil M., *Risk Assessment and Decision Analysis with Bayesian Networks*, London: CRC Press, 2019, p. xvii.

A Bayesian Artificial Intelligence Model for Legal Fact-Finding

Liu Hai

(School of Statistics and Information, Shanghai University of International Business and Economics,
Shanghai 201620, P.R.China)

Abstract: Accurate fact-finding is the premise and foundation of judicial fairness and justice. In judicial practice, the principle of discretionary evaluation of evidence is the proof system for fact-finding. However, discretionary evaluation of evidence faces a triple dilemma: difficulty in clarifying evidence, implicit subjective judgment, and fuzzy inner conviction, which hinder the accurate determination of the truth. The Bayesian Artificial Intelligence (AI) model, as a visual, transparent, and quantitative intelligent tool, can assist fact-finders in completing the process of discretionary evaluation, uncovering the black box of discretionary evaluation, and promoting the accuracy of fact-finding. Bayesian AI uses Bayesian network diagrams as the knowledge system to visually represent the inferential relationships between evidence and facts. It utilizes probability numbers as the data system, transparently displaying the fact-finder's subjective judgment on the degree of dependence between evidence and facts. Probabilistic algorithms are used as the inference system to intelligently calculate posterior probabilities, quantifying the strength of inner conviction in inference. The Bayesian AI model offers five advantages:

First, Bayesian AI can not only use machine learning methods to process big data but also integrate smart data such as expert knowledge and judgments. When dealing with risky decisions with little or no data, such as court trials, Bayesian AI can combine human knowledge and experience to provide viable analytical models.

Second, Bayesian network inference makes the propagation and updating of the inferential strength of evidence in continuous inference logical and scientific. It can not only handle corroborating evidence but also uniformly analyze conflicting evidence.

Third, the application of software tools solves the dilemma of “weak mathematical calculation” faced by fact-finders and promotes the application of the Bayesian AI model in the practice of fact-finding. Fact-finders who are not proficient in probabilistic reasoning can effectively use software tools to construct Bayesian network models after simple training.

Fourth, the Bayesian AI model helps to open the black box of discretionary evaluation, presenting the subjective judgment and reasoning process of fact-finders in a visual, transparent, and quantitative manner, thereby promoting the accuracy of fact-finding.

Fifth, Bayesian AI is a probabilistic reasoning expert system based on human-computer collaboration. Its goal is to “assist people” and “enhance people” rather than “replace people.” All judicial decisions are still made by humans, and the subject status of fact finders remains unchanged.

However, the Bayesian AI model also faces challenges. First, it requires a large number of probability numbers as input, but fact-finders' subjective judgments on evidence are difficult to accurately transform into probability numbers, thus affecting the accuracy of the model's outputs to a certain extent. Second, while the Bayesian AI model can avoid logical fallacies in evidential reasoning and is a tool for the explicit expression of subjectivity, it cannot eliminate subjectivity.

Keywords: Discretionary evaluation of evidence; Bayesian artificial intelligence; Fact-finding

[责任编辑:王玲强]