刑事司法科学与治理
Criminal Justice Science & Governance
2020 年 第 1 卷 第 2 期

手机取证数据在毒品案件线索挖掘中的应用

刘元生

福建中锐电子科技有限公司，福州

摘　要 | 毒品类案件危害大，线索来源少，而手机取证数据是涉毒案件侦查的重要证据，本文提出针对手机取证数据，建立涉毒情报线索挖掘模型，固化毒品案件侦查技战法，应用知识图谱等大数据分析技术，构建涉毒群体虚拟身份关系网络，分析挖掘潜在的涉毒人员虚拟身份线索，实现从数据到情报的转换，为毒品案件侦查提供有价值的情报。

关键词 | 知识图谱；涉毒线索挖掘；手机数据分析

根据国家禁毒委员会办公室发布的《2019 年中国毒品形势报告》指出，毒品贩运活动依然活跃，呈现境内境外、网上网下相互交织的局面。吸毒方式越来越隐蔽，排查发现难。为规避公安机关查处，吸毒活动隐蔽性私密性特点增强，公共娱乐场所吸毒活动有所减少，选择在宾馆、出租屋、私人会所或私家车等隐蔽处所吸毒明显增多；一些吸毒人员从线下转入线上，利用网络社交软件建立“毒友群”，采用虚拟身份、暗语交流，进群先直播吸毒，进群后不参与直播吸毒或不购买毒品即被踢出群，形成更加隐蔽的网络吸毒圈子[1]。

一、毒品类案件的特点

（一）毒品案件危害大、影响大

第一是对家庭危害大。吸毒者在自我毁灭的同时，也破坏着自己的家庭，使家庭陷入经济破产、妻离子散、家破人亡的困境；第二是毒品活动严重扰乱社会治安。吸毒者吸食、注射毒品，需要大量的金钱，吸毒者面对高额的费用和强烈的诱惑，会丧心病狂、不择手段、甚至铤而走险进行抢劫、盗窃、诈骗、贪污、卖淫甚至杀人等违法犯罪活动，给社会治安造成严重危害。大量事实证明，吸毒已

［1］2019 年中国毒品形势报告［EB/OL］．（2020–06–24）．http://www.nncc626.com/2020–06/24/c_1210675813.htm.

成为诱发犯罪、危害社会治安的根源之一。

（二）涉毒线索来源渠道少、获取线索成本高[1]

毒品案件与其他普通刑事案件相比，有其特殊性。一是无典型意义上的被害人，不会主动报案。因为有的毒品案件不直接侵害到人，如种植、制造毒品行为，一般无被害人报案。毒品案件交易双方都是不法分子，不会有人报案。二是毒品案件一般无典型意义上的现场。如走私和贩卖毒品案件，买卖双方钱货两清，犯罪分子不会在交易地点留下痕迹物证，通过现场勘查取证难。三是毒品案件知情人少。由于毒品案件隐蔽性的特点，犯罪行为诡秘，群众很难知晓内情，通过群众报案或调查访问获取线索不可行，线索来源十分缺乏。

（三）搜集和固定证据难

1. 毒品容易灭失，搜集认定难。缉毒部门在查缉毒品案件，毒贩们一旦发觉贩毒行为暴露，会迅速销毁毒品，导致出现因证据不足而难以定罪的情况。即使人赃俱获，也很难查清毒品源头和去向。甚至还有毒品被吸毒人员消费而难以查清毒品的数量，因而妨碍到对毒贩的定罪量刑。

2. 证人证言查找难。毒品案件主要有三类证人：第一类证人是吸毒人员，因其特殊的人格特征，证言的可信度不高；第二类证人在毒贩欺骗下不明真相为其走私运输毒品的人，一般不愿意与警方合作，同时因为毒贩行动诡秘，不用真实的姓名、住址，流动性大，因此对交易内幕难以了解；第三类证人是目睹毒贩从事贩毒活动的其他人，因为毒品案件隐蔽性强，主要在圈内交易，圈外人很难获取有价值的信息，因此侦查人员想通过他们获取证人证言，更是难上加难。

3. 证据转化难。毒品案件侦查多用秘密跟踪、秘密搜查、技侦手段监控等秘密措施，获得的材料有时只能作为提供侦查方向的依据，无法转换为证据使用。

4. 获取毒品犯罪嫌疑人的口供难。我国目前的刑事法律对毒品犯罪的处罚普遍较重，因此犯罪嫌疑人被抓获后，受畏罪心理支配，知其涉毒数量大，交待了可能会被判以重刑；有的受侥幸心理支配，了解缉毒部门案件线索来源少、证据搜集难等状况，企图蒙混过关，拒不交待。因此，缉毒部门获取毒品犯罪嫌疑人的口供很难。

（四）手机取证数据的价值未充分利用

智能手机已经广泛应用于人们的日常生活，在涉毒案件中，智能手机中的通联行为、轨迹信息、消息通联、资金交易等是证据的重要来源途径。因此，手机取证数据在毒品案件中得到广泛的使用，但依靠侦查人员人工查阅手机取证数据会存在一定局限性，大量有价值的情报线索未得到挖掘和利用，主要有以下方面因素。

1. 取证数据量大。手机取证数据中包含大量有价值的数据，包括电话联系人、电话通联记录、出行数据、资金交易、消息通联等，但关键证据淹没在大量的无效信息中，需要通过分析类软件能快速挖掘和检索到关键的线索片段。

2. 犯罪嫌疑人的关系网络错综复杂。移动智能终端上的通话、短信、即时通信、微博、地理信息、电子商务等应用数据之间存在关联关系，如何在复杂的关系网络中快速定位犯罪线索路径，需要分析软件支撑。

3. 涉毒群体的本地化特征和网络化特征显著。涉毒群体具有延续性、关联性和本地化特征，涉毒群体呈现出一个封闭的犯罪网络，这需要缉毒部门构建本地化涉毒嫌疑人电子取证数据资源库，可以利用历史案件数据碰撞和关联分析、挖掘出新案情、新线索，从而理清本地涉毒的网络组织结构。

基于以上背景，如何利用大数据技术分析和挖掘嫌疑人手机中包含的犯罪线索是非常有价值的课题，具有重大意义。本课题基于手机取证数据，通过建立涉毒情报线索挖掘模型，利用已有涉毒人员的手机取证数据，分析挖掘出潜在的涉毒人员虚拟身份线索，帮助警方理清本地的涉毒网络，确定涉毒的上线和源头，有效打击毒品类案件，营造健康、有序、和谐的本地治安环境。

二、应用研究现状

在万物互联时代，以电信诈骗、网络诈骗、P2P、网络赌博、网络贩毒等新型违法犯罪活动，已经成为公安机关案件侦查重点，移动智能终端是

[1] 许翠华．毒品案件侦查中的问题探析［J］．江苏警官学院学报，2004，19（2）：33–35.

实施犯罪的重要载体，终端上的电子数据，是侦破案件的重要证据来源。公安部为应对新形式下案件特点，2018年6月25日，公安部第五局关于印发《电子物证检验实验室建设规范》的通知，明确要求各地公安规范电子物证检验实验室建设，进一步提升电子物证建设水平，充分发挥电子物证支撑侦查破案的重要作用。各地公安陆续建设了电子取证实验室，手机取证设备及其数据分析已经广泛应用于公安案件侦查研判，各地公安电子取证实验室累积了大量的本地化涉案电子取证数据，如何有效利用本地积累的涉案电子取证数据，是各地公安面临的一个重要课题。

针对本地化涉案电子取证数据的利用，各地公安做了大量有价值的实践，主要是在省级集中资源在建设各类大数据分析平台，例如美亚蛛网、烽火公安大数据、明略科技小明等，都是在全国或省一级建设的公安大数据分析平台，从更大的维度和视角，汇聚各类涉案电子数据、各类视图数据、社会数据以及公安业务数据，对案件侦查研判起到了非常重要的作用。但是在针对某个具体案件类型和本地化案件特征，缺乏针对性、专业化的深入挖掘分析，没有最大化挖掘出数据的价值。例如针对涉毒类案件，涉毒群体的本地化特征和网络化特征非常显著，可以根据本地已知涉毒群体的手机取证数据，进行侦查扩线，理清毒品交易的上线和下线，通过构建本地化涉毒人员电子取证大数据，利用大数据分析建模方法，实现案件串并研判，挖掘出新案情、新线索，帮助理清本地涉毒网络组织结构。

在针对涉案电子取证数据的分析技术方面，知识图谱是公安大数据应用的主流方向。所谓知识图谱，就是在大数据分析的基础上，通过语义理解将“点线面”的数据关联与事物现实中非简单指向性的复杂关系相联结，从而形成实用型认知应用。知识图谱主要解决的，就是公安多年积累的实战经验与技术算法如何相互转换的问题。国务院印发的《新一代人工智能发展规划》中，明确提出要构建覆盖数亿级的知识实体的多元、多学科、多数据源的知识图谱。对此，公安部第一研究所信息技术事业部副主任汪宁表示，这表明知识图谱是国家在人工智能领域布局的重要方向之一。2018年9月7日，公安部第一研究所牵头、明略数据联合编写的业内首个《公安知识图谱标准与白皮书》正式发布，标志知识图谱将成为公安大数据应用的主流方向之一，建设针对具体业务的知识图谱实战解决方案将成为行业焦点。在针对涉毒人员手机取证数据线索挖掘应用方面，知识图谱等大数据分析技术，具备先天的优势，因为涉毒群体本身是一个封闭的网络，且本地化特征明显，因此应用知识图谱技术从涉毒手机取证数据进行线索挖掘，是技术与应用结合的创新。

基于目前公安手机取证数据的应用现状，本文认为，针对公安具体业务的数据分析研判模型和线索挖掘，是一个重要的研究课题。特别是针对涉毒类案件，线索来源少，获取线索的成本高代价大，而涉毒案件的本地化特征和网络化特征又非常显著，各区县公安累积的大量涉毒人员手机取证电子数据，是非常宝贵的财富，应用信息化手段和大数据分析技术，结合涉毒案件侦查技战法，为涉毒案件侦查提供有价值的线索，意义重大且非常有实战应用价值。

三、问题和难点

智能手机中保留生活、工作、社交、娱乐、出行、购物、支付等大量数据，如何把杂乱、无序的非结构化数据转换为有序的结构化数据，结合毒品案件侦查技战法，利用大数据和人工智能技术，分析和挖掘涉毒案件取证数据中未抓获的涉毒人员虚拟身份线索，是本文要解决的重点问题，但在利用手机取证数据进行线索分析和挖掘时，面临的主要问题和难点包括以下方面。

（一）手机取证获取的数据有限，存在数据缺失、数据不完整等问题

手机取证获取的数据有限，存在数据缺失、数据不完整等问题导致数据割裂，这严重了影响线索挖掘建模分析。智能手机中的应用数据一部分保存在本地，另一部分保存在远程服务器。因此，缉毒部门要想获得智能手机中的应用数据只能提取和恢复手机上保存的数据。同时，由于技术的局限性，现有的手机取证设备提取和恢复的手机数据存在数据不完整、数据缺失等问题，会影响分析结果。

（二）大量有价值的证据数据包含在非结构化的数据中

大量有价值的证据数据包含在非结构化的数据

中，例如短信中的出行订票信息、银行交易通知、微信消息 / 拍照中的位置信息、银行卡 / 身份证西等涉案证据信息。这需要缉毒部门利用数据清洗、模板匹配和自然语言处理等技术，将有侦查分析价值的数据进行结构化。

（三）犯罪通联方式多样化

嫌疑人在利用手机组织和实施犯罪时，存在多种通联方式和虚拟身份，各种虚拟身份之间是割裂的，例如同一个嫌疑人存在多个电话号码和微信账号，沟通方式有通话、短信、微信 /QQ、微博等，如何有效关联不同虚拟身份和通联方式，是数据分析挖掘的难点。除此之外，手机中包含了大量与案件无关的数据，如何从杂乱、无序的数据中区分有价值的信息，也是数据分析挖掘的难点。

（四）侦查人员对业务规则缺乏准确和量化的表达，很难转换为计算机可理解的业务规则

毒品案件侦查技战法是办案人员的经验，提炼和总结毒品案件侦查技战法的业务规则本身是一个很难的过程，侦查人员很多的时候是依靠经验和感觉，缺乏准确和量化的表达，很难转换为计算机可理解的业务规则，例如侦查人员反馈说来自毒源地的虚拟身份，在实际办案中是一条有价值的重要线索，但理清这条线索挖掘规则，需要梳理出一系列的判断准则，如是否和多个贩毒人员存在关联、是否和毒贩存在毒品相关的通联行为、是否存在资金上的往来、是否是去毒源地旅游度假等等。

（五）毒品案件中证据数据的真实性是否有效的验证

毒品案件线索挖掘中，判断一条毒品案件线索是否真实有效，存在较多的前提条件，而验证这些前提条件本身也是一个难点，既存在数据本身的难点，也存在技术本身的难点。例如毒贩在微信通联时，经常使用“东西”作为毒品暗语，但判断包含“东西”的通联行为是否毒品交易，需要对通联上下文进行判断，需要理解双方沟通的语义，技术实现本身存在难点。

四、建设思路

在手机取证数据中挖掘毒品线索，本质上是将数据转换为情报的过程，可以用图 1 形象的表达[1]。

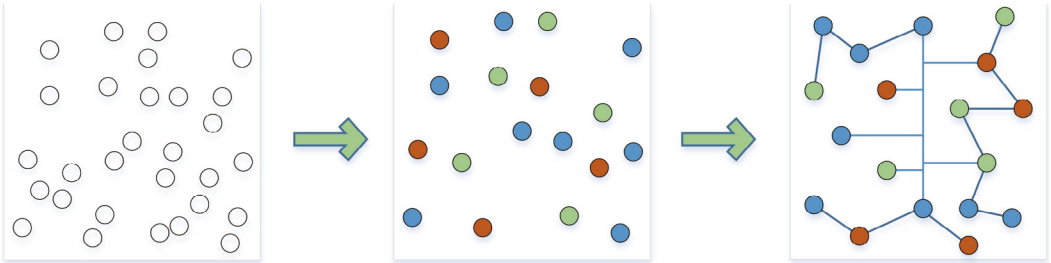

图 1　数据到知识转换关系图

手机取证数据挖掘毒品线索可以细化为以下几个主要的过程：

（一）线索挖掘建模

手机取证数据挖掘毒品线索，本质是将毒品案件的侦查技战法转换为线索挖掘模型。这是基于历史涉毒案件的电子取证数据，并结合涉毒案件特征、涉毒人员关系网络、涉毒人员行为习惯、资金交易特点等维度进行的业务抽象建模，总结和提取线索推理的业务规则，构建涉毒情报线索挖掘模型，整体建模思路如图 2 所示。

[1] 知识图谱标准化白皮书（2019）[EB/OL]. [2020-09-28]. http://www.cesi.ac.cn/201909/5589.html.

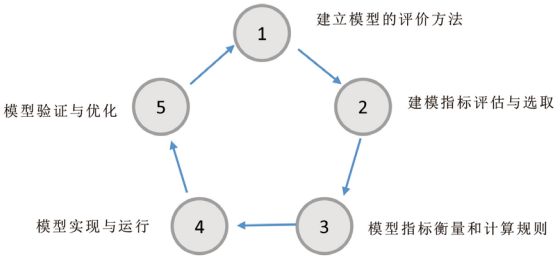

图 2　建模思路

为了评价模型的有效性，首先需要建立模型有效性的评价方法，本文中采用以下规则评价模型的有效性：

1．对模型指标评分，采用百分制来衡量线索的有效性，分值越大，线索的有效性越高；

2．模型由若干指标组成，每个指标有对应的权重，权重用于衡量指标的重要程度，指标权重越大，说明该指标在模型越重要，每个指标有命中分值，指标命中分值和权重的乘积得到指标的最终得分，最后汇聚模型中所有指标的得分，得到线索的分值；

3．对模型命中结果，最终统一使用可信度来评价，采用分值区间衡量线索的可信度，划分为 P1 到 P5 共 5 个等级，P5 对应的线索有效性最高，P1 对应的线索有效性最低。

建模指标评估与选取是整个建模过程的关键，是毒品案件侦查技战法的总结，结合手机取证数据，本文从图 3 所示的 8 个维度指标进行建模分析。

图 3　涉毒线索挖掘指标建模

由于篇幅所限，笔者在此仅重点介绍涉毒人员关联关系指标，其它指标的建模方式与此类似。由于毒品交易的网络化特征非常明显，毒品交易是一个封闭生态圈，涉毒群体存在延续性、关联性等特征，通过分析本地历史涉毒案件电子取证数据，可以找出多起涉毒案件中关联的联系人。因此，通过涉毒人员关联关系可以挖掘出潜在的涉毒人员。涉毒人员关联关系指标是指被评估对象与多少个涉毒人员相关，关联方式可以包括电话通讯录、通话、短信、彩信、即时通讯好友、即时通讯消息通联、资金交易等方式，被评估对象关联的涉毒人员数量越多，参与涉毒的概率越大[1]。假设 B 表示被评估对象，R（B）表示关联涉毒人员数量，S 表示关联 1 个涉毒人员的分值，W 表示指标权重，则被评估对象 B 的该指标最终分值 =R（B）* S * W。

图 4 是某涉毒案件应用涉毒人员关联关系指标分析挖掘的可视化结果呈现，该被评估对象关联到 8 个涉毒人员、7 起涉毒案件，涉毒的本地化网络特征非常显著。

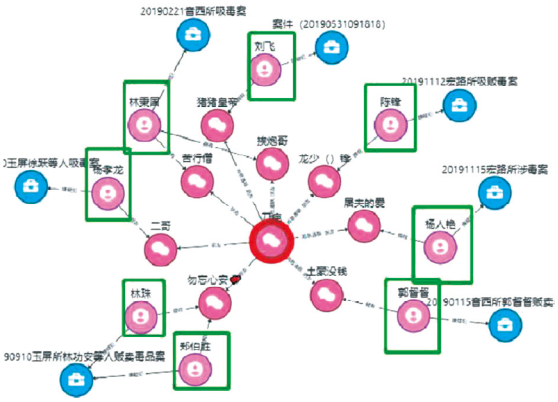

图 4　涉毒人员关联关系案例

（二）数据清洗

智能手机已经是犯罪重要的实施工具和载体，手机取证数据包含涉案嫌疑人各方面的数据，犯罪线索隐藏在大量杂乱无序的原始取证数据中，实现线索挖掘的基础需要对原始的取证数据进行结构化，将数据清洗和预处理为线索挖掘所需的结构化数据，建立数据之间的关联关系。

目前主流的手机取证产品包括厦门美亚柏科的 DC4501、大量睿海的 RH–6900 和四川效率源的 SPF9139 等，不同手机取证产品各有优劣，基本上都能支持手机上主流应用的手机数据提取和恢复，

［1］王刚，刘猜．犯罪团伙网络关系模型及分析方法［J］．中国人民公安大学学报，2015（3）：61–71.

为了能利用和分析各厂商的手机取证数据，公安机关制订了 GA/WA 1006–2013《网安数据传输交换规范》、GA/WA 2003–2019《电子数据取证设备技术规范》等一些列数据交换的规范，用于实现取证设备输出数据与分析类系统数据互通，因此在数据清洗阶段，可以遵循公安部门制订的规范和标准，根据业务建模的需要，对手机取证数据进行清洗和结构化，建立业务领域专题数据库[1]。

（三）涉毒情报模型实现

涉毒情报模型实现，就是使用编程语言和数据挖掘算法实现和固化侦查技战法，通过数据清洗技术，构建涉案人员电子取证数据资源库，应用人工智能技术、知识图谱技术和数据挖掘算法，构建嫌疑人物理和虚拟身份关系网络，从时间、空间、社交、资金等多维度实现案件分析研判，全方位重构涉案人物画像和关系网络，应用涉毒情报线索挖掘模型，根据涉毒人员关系网络、资金往来、通联行为、时序关系、暗语特征、活动轨迹等多维度指标，从手机取证数据中挖掘出潜在的涉案情报线索。在涉毒情报模型实现中应用了非结构化数据库技术、知识图谱技术、全文检索技术、数据挖掘算法等，在此重点介绍知识图谱技术在毒品线索挖掘中的应用。

涉毒情报模型的本质是梳理毒品交易的关系网络，发现毒品交易的关系网络就是通过建立各类电子证据之间的关联关系，而实现电子证据关联关系最直接的技术方案就是知识图谱。知识图谱本质就是一张图，这张图把每个知识点作为一个个节点，知识点之间的关系作为连接这些节点的边，通过一张图串联所有线索。知识图谱，是用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。知识图谱是一种用于表达关系的网络，通常用“实体(Entity)”来表达图里的节点、用“关系(Relation)”来表达图里的“边”。知识图谱非常适合用于电子数据之间的关联性分析，通过知识图谱可以串联不同类型的电子数据，例如电话号码之间的通联关系、电话号码与微信账号的绑定关系、微信账号之间的朋友关系、嫌疑人与案件之间的涉案关系等，同时应用分析模型算法，可以挖掘虚拟身份之间的深层次关系，进一步丰富知识图谱的表达，实现的知识演化，图 5 是知识图谱技术的形象表达。

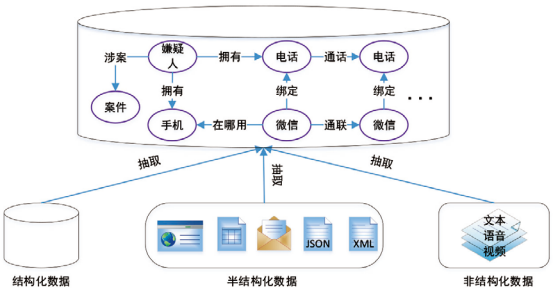

图 5　知识获取示意图

知识图谱的应用，关键在于建立知识图谱的业务领域模型，归纳为以下几个要点[2]：

1. 发现和识别嫌疑人手机取证数据中的实体和关联关系。这个属于基础层面的关联关系，对嫌疑人取证数据中的通讯录、通话记录、短信、微信、QQ、微博、支付宝等数据，建立对应的实体及关联关系。

2. 分析业务领域关注的实体和关联关系。这需要使用数据清洗、模板匹配、自然语言处理技术等从手机数据中提取案件关键要素并建立关联关系，例如从银行交易通知短信中提取交易账号及其关联关系，从微信消息中提取涉案文档进而发现微信账号之间的涉案关系。

3. 识别和挖掘深层次关联关系。根据线索模型和分析算法，挖掘有价值的情报和线索。实体之间最基础的关联关系，例如两个电话号码之间的通联关系，对于侦查人员没有直接利用价值，以涉毒案件为例，如果发现某个电话号码与多个涉毒人员存在关联，同时发现该电话号码在通话记录中符合伴随关系或时序关系的特征，则该电话号码涉毒的概率较大，对于侦查人员而言，这是一条有价值的情报。

4. 建立数据标签，固化已有的知识。缉毒部门将已有发现的线索、关联关系和数据特征，转换

[1] 李双其，林伟. 侦查中电子数据取证 [M]. 北京：知识产权出版社，2018：166–201.

[2] 刘元生. 知识图谱在手机数据线索挖掘中的应用 [J]. 福建电脑，2020，36（10）：95–97.

为可分析利用的数据标签，实现知识的演化和升级。例如将涉案虚拟身份标注涉案类型标签，将分析挖掘得到的团伙关系、借贷关系进行标注，从而丰富知识内涵，这些知识又有利于进一步挖掘新的情报。

五、应用

在福州市禁毒支队大力支持下，本文利用涉毒情报线索挖掘模型对福清市禁毒大队近两年涉毒案件进行分析。其中，包含共 128 个案件，390 个嫌疑人，541 部手机，将历史涉毒案件手机取证数据导入到 E 探电子数据分析系统，使用涉毒情报线索挖掘模型挖掘潜在的涉毒人员线索，最终得到可信度为 P4 和 P5 的线索共 256 条。提供的线索由于都是应用软件的虚拟身份账号，所以无法直接确定人员身份。人员身份的确定又依赖于公安部门技侦手段和工具以及核实的成本比较高。因此，笔者最终从 256 条线索中任意挑选了 59 条线索进行身份落地，其中 9 条线索身份无法落地，1 条线索无效，共落地 49 个嫌疑人，其中 13 个人是已经被批捕的涉毒人员，剩余 36 个落地身份是未批捕的涉毒人员，但需要持续经营，分析结果有效性得到福州市禁毒支队肯定。

六、结束语

在信息化时代，手机已经成为重要的作案工具和数据载体。手机取证数据是案件侦查的重要线索来源，特别是在涉毒案件中，手机取证数据大有可为。目前，手机取证数据在执法单位的应用更多处于辅助性质，局限在案件本身，手机取证数据价值还没有得到充分利用。从手机取证数据挖掘毒品案件线索，关键在于数据挖掘算法，而数据挖掘算法规则来自公安一线侦查人员，需要根据案件类型总结和提炼相应的技战法模型。针对手机取证数据进行建模分析，侦查人员需要提升数据挖掘能力和线索的有效性实现知识演化和升级。构建和积累本地化的涉毒犯罪群体中的手机取证大数据，为涉毒案件侦查提供有价值的情报和分析研判支撑。

（责任编辑：田　伟）

Application of Phone Data Clue Mining in Drug Crime Case

Liu Yuansheng

Fujian Zhongrui Electronic Technology Co., Ltd, Fuzhou

Abstract: Drug-related cases have great harm and hard to get clues. Phone data is an important evidence in drug-related case investigation. This paper base on suspect's phone data, introduces how to establish a drug-related intelligence clue mining model, apply big data analysis technology such as knowledge map, construct virtual identity relationship network of drug-related groups, analyze and mine potential drug-related clues, realize the conversion from data to clues, and provide valuable clues for drug-related case investigation.

Key words: Knowledge map; Drug-related clue mining; Mobile data analysis